SSU72 AND RTR1 SERINE 5 PHOSPHATASES AND THEIR ROLE IN NNS AND CPF TRANSCRIPTION TERMINATION

Jose Fabian Victorino

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Biochemistry and Molecular Biology,
Indiana University

May 2020

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____
Amber Mosley, Ph.D., Chair.

_____
Peter Roach, Ph.D.

January 31, 2020

_____
Millie Georgiadis, Ph.D.

_____
Yunlong Liu, Ph.D.

_____
Gustavo Arrizabalaga, Ph.D.

DEDICATION

This work is dedicated to my family, who have given me the tools that I need to accomplish anything I can conjure. My mother, Arcelia Victorino-Lopez, who has always supported me whether it was my head as a newborn infant or my goals as a man, there is no son who could have a better mother. My father, Jose Victorino Sr., who has and continues to inspire me to push my limits further than anyone could ever reach. My sister, Ana Victorino, who provided a highly competitive childhood through which I was able to expand my abilities and strengthen my resolve. My love, my muse, my nucleus, Loan Anh Thi Do, who has helped shoulder all of my dreams, who picks me back up when they become too heavy and I grow too weary, and who fuels my passion when it is low.

ACKNOWLEDGEMENT

I would like to thank Indiana University School of Medicine for allowing me to attend its great institution. Through this program I have grown a great deal personally, professionally and academically. I would like to thank my committee members Doctors Amber Mosley, Peter Roach, Millie Georgiadis, Gustavo Arrizabalaga, and Yunlong Liu for their invaluable support and advice throughout my stay at IUSM and in writing this thesis. These professors are outstanding mentors and have provided an awesome environment for me to grow professionally and personally.

I would also like to thank the Mosley team, without them the success I have had would have been crippled. In particular I would like to thank those who have trained me in experiment and theory; Dr. Amber Mosley, Dr. Melanie Fox and Whitney Smith-Kinnaman. I also owe a special thanks to Dr. Menghao Hoang, Sarah Peck Justice, Katlyn Hughes Buriss, Dr. Jason True, and Dominique Baldwin who have helped me tremendously through the sharing of ideas, methodology, general advice, written discourse, random discourse, and some good old times. In addition, I am grateful to have shared the lab floor with several great individuals who have frequented the lab during my stay, several of whom have helped me throughout my studies; Dr. Lynn Bedard, Asha Boyd, Gabi Mazur, Samuel Ogunsanya, Rachel Chan and Neil McCracken.

I also must thank the Biochemistry department for all of the support and equipment that they have provided in particular the Dong lab, the Georgiadis lab and the Biochemistry staff.

Jose Fabian Victorino

SSU72 AND RTR1 SERINE 5 PHOSPHATASES AND THEIR ROLE IN NNS AND CPF TRANSCRIPTION TERMINATION

Polyadenylation dependent transcription termination is dependent on the Cleavage and Polyadenylation Factor complex (CPF) which is essential for the termination and processing of mature RNA. Polyadenylation (PolyA) independent transcription termination is carried out by the NNS (Nrd1-Nab3-Sen1) termination pathway, which helps regulate termination and processing of non-coding RNA (ncRNA). The disruption of these pathways can impact expression of nearby genes, both protein coding and non-coding. Recruitment of termination pathway components is achieved through a domain unique to the largest subunit of RNA Polymerase II (RNAPII) referred to as the C-terminal domain (CTD), which contains a repeating heptad sequence, $Y_1S_2P_3T_4S_5P_6S_7$, and acts as a docking site for transcription regulatory proteins. Ssu72 is a serine 5 phosphatase and an essential member of the CPF complex. Rtr1 is also a serine 5 phosphatase, but its mechanism of action is less well characterized. Both Rtr1 and Ssu72 regulate transcription machinery recruitment through control of the phosphorylation status of the CTD. My studies have focused on Rtr1 and Ssu72 mutants in yeast which show evidence of transcription termination related phenotypes. Chromatin immunoprecipitation of RNAPII followed by exonuclease treatment (ChIP-exo) studies provide evidence of RNAPII transcription continuing through termination sites at ncRNA genes as a result of a hyperactive Ssu72-L84F mutant, while an *RTR1* knockout results in increased premature RNAPII transcription termination. Northern blots and RNA sequencing confirm premature transcription termination and decreased total RNA

expression in the *RTR1* knockout and increased length of ncRNA transcripts as well as total RNA expression in the Ssu72-L84F mutant. Mass spectrometry analysis has identified changes in the protein-protein interactions (PPI) within the CPF complex in the Ssu72-L84F mutant and decreased PPIs between different transcription machinery in *RTR1* knockout cells. My results show that the CTD phosphatases Rtr1 and Ssu72 play unique roles in the regulation of RNAPII termination in eukaryotes.

Amber Mosley Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4tU-seq	4-thiouracil labeling with RNA sequencing
ALS4	Amyotrophic lateral sclerosis type 4
AOA2	Ataxia oculomotor apraxia type 2
APA	Alternative polyadenylation
AS	Antisense
ATP-32P	Adenosine triphosphate labeled with radioactive 32P
bp	Base pair
CBC	Cap binding complex
CF	Cleavage factor
ChIP	Chromatin immunoprecipitation
ChIP-chip	Chromatin immunoprecipitation on microarray
ChIP-exo	Chromatin immunoprecipitation with exonuclease digestion
ChIP-seq	Chromatin immunoprecipitation with DNA sequencing
CID	CTD interacting domain
CML	Chronic myelogenous leukemia
CPF	Cleavage and polyadenylation factor
CRAC	Crosslinking and analysis of cDNAs
CTD	C-terminal domain
CTDK1	CTD kinase 1
CUTs	Cryptic unstable transcripts
DisCo	Disruption compensation network
DNA	deoxyribonucleic acid

EDTA	Ethylenediaminetetraacetic acid
ET	Extended transcript
FDR	False discovery rate
lncRNA	Long noncoding RNA
mRNA	messenger RNA
MUTs	Meiotic unstable transcripts
NNS	Nrd1-Nab3-Sen1 termination machinery
nt	Nucleotide
NUTs	Nrd1-unterminated transcripts
OD600	Optical density measured at a wavelength of 600 nm
ORF-T	Open reading frame transcript
P#	Preparation number
PAR-CLIP	Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation
PCH1	Pontocerebellar hypoplasia type 1
PCR	Polymerase chain reaction
PNK	Polynucleotide Kinase
RIN	RNA integrity number
RNA	ribonucleic acid
RNAPII	RNA Polymerase II
RRM	RNA recognition motif
rRNA	ribosomal RNA
SDS	Sodium dodecyl sulfate

S2P	Serine 2 phosphorylation
S5P	Serine 5 phosphorylation
S7P	Serine 7 phosphorylation
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SUTs	Stable unannotated transcripts
T4P	Threonine 4 phosphorylation
TFIIH	Transcription factor II H
TOV	Terminator override
ts	Temperature sensitive
TSS	Transcription start site
TTS	Transcription termination site
Y1P	Tyrosine 1 phosphorylation
USE	Upstream Sequence Element
UTR	Untranslated region
WT	Wild type
XUTs	Xrn1-unterminated transcripts
YPD	Yeast extract-peptone-dextrose medium

INTRODUCTION

RNA Transcription

The process of transcription is highly regulated in order to achieve proper differential gene expression, and yet full understanding of the mechanisms involved in its initiation, elongation, and termination are still lacking. With the advent of new technology and techniques, gains have been made in understanding transcription and discovering more direct relationships between the transcriptional machinery and human disease. Proteins involved in transcription have been linked to a wide range of diseases, such as human immunodeficiency virus (HIV), neurodegenerative ataxias, ataxia with oculomotor apraxia type 2 (AOA2), and myofibrillar myopathies [1-7]. Many cancers have been discovered to be "transcription addicted" meaning that they require the upregulation of transcriptional machinery in order to proliferate [8]. These observations underscore the need for high levels of regulation during transcription, and that all players in the process of transcription are very important to achieve a working system that can receive nuclear, cellular, and environmental signals to regulate appropriate expression of RNA.

1. RNA Polymerases

In eukaryotes there are five RNA Polymerases (RNAPI-V), each polymerase with a specialized role in the production of specific subclasses of RNA. RNAPIV and RNAPV are only found in plants and will not be discussed at length here. RNAPI, RNAPII, and RNAPIII have a conserved catalytic core of 10 proteins Rpb5, Rpb6, Rpb8, Rpb10,

Rpb12 and five other subunits that are encoded by different genes depending on the which polymerase (RNAPI/RNAPII/RNAPIII) they are a part of; A190/Rpb1/C160, A135/Rpb2/C128, AC40/Rpb3/AC40, AC19/Rpb11/AC19 and A12.2(N ribbon)/Rpb9/C11(N ribbon) [9]. The three polymerases also have homologous heterodimeric subcomplexes A14/43, Rpb4/7, and C37/53 that form a stalk for RNAPI, RNAPII, and RNAPIII respectively [10]. RNAPI transcribes the ribosomal 18S, 5.8S, and 28S RNA (rRNA) which represents ~80% of the total RNA produced in a yeast cell [11]. RNAPI contains 14 subunits, there are two proteins in this complex that are not homologous to RNAPII being A49 (N-terminal domain) and A34.5 that have homology to subunits of transcription initiation factor F (TFIIF) [10]. RNAPIII synthesizes transfer RNA (tRNA) and 5S rRNA that account for 15% of total RNA [11]. RNAPIII contains 17 subunits; the additional proteins that are not homologous to RNAPII are C37, C53, C82, C34, and C31 [10].

RNA polymerase II (RNAPII) transcribes messenger RNA (mRNA) as well as a diverse set of other transcripts including: microRNA (miRNA), long non-coding RNA (lncRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), and other non-coding RNA (ncRNA) (Figure 1) [12, 13]. RNAPII is thought to be responsible for cellular differentiation as well as to maintain the identity of a cell when responding to environmental signals.

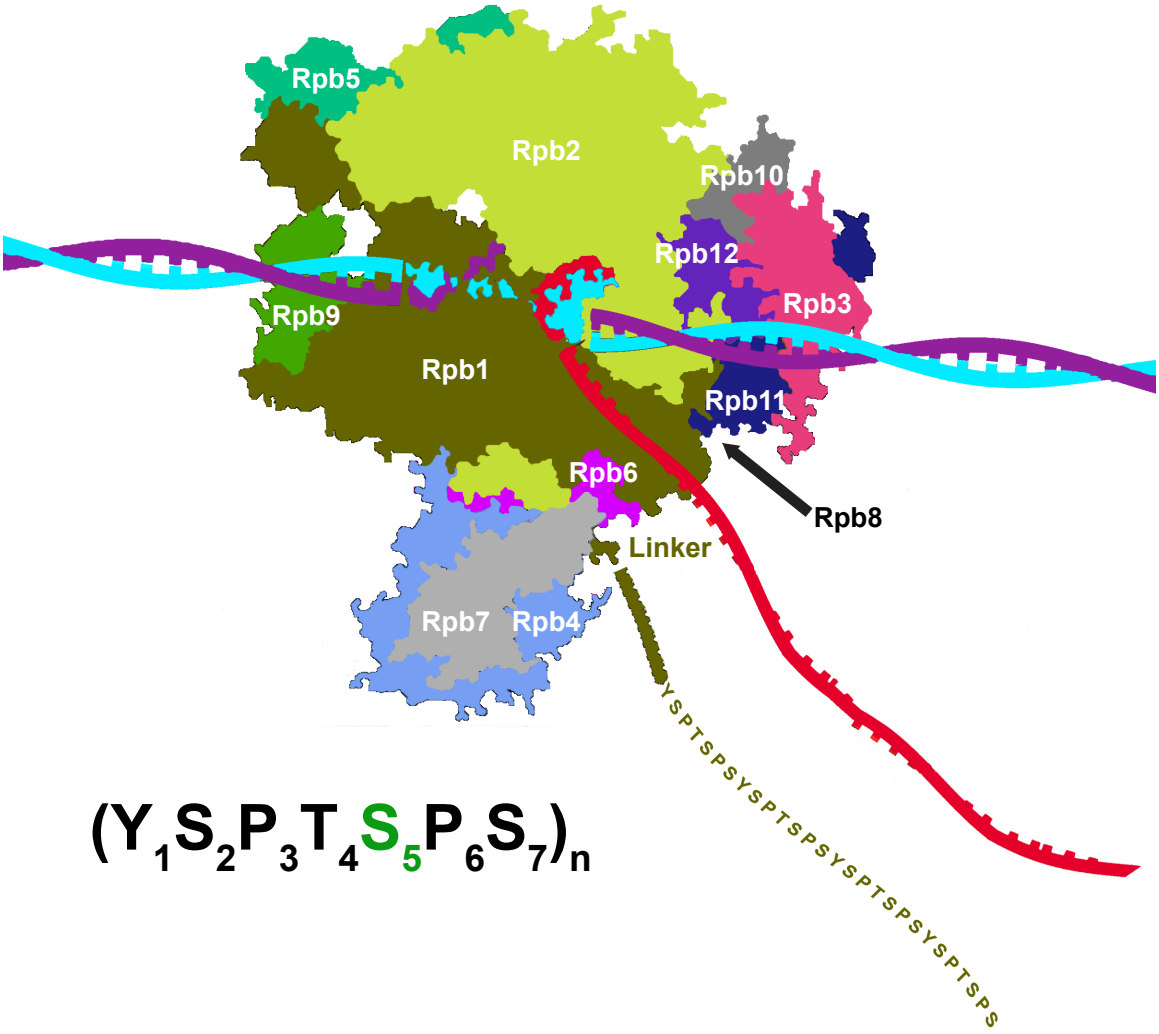


Figure 1: RNA Polymerase II complex. Artistic rendering of a structure of a paused RNAPII complex (PDB ID: 3HOU) [14]. RNA is in red; DNA strands are in purple and light blue. The CTD is truncated at the linker region before heptad repeats begin. Serine 5 is highlighted for its significance to this work.

RNAPII is a 12-subunit complex with a catalytic core of ten proteins (Rpb1-3, 5,6,8-12) and two dissociable subunits Rpb4 and Rpb7 that appear to be most associated with the RNAPII core during initiation with interactions decreasing during continuing elongation. [15, 16]. Rpb1 is the largest of the core proteins at ~220kDa and contains a regulatory domain, called the C-terminal domain (CTD) with a sequence of heptad amino acid repeats (YSPTSPS), which directs the different stages of transcription from initiation to elongation [17]. The other RNAPII subunits have several functions which help grip DNA/RNA and form different channels for these molecules to enter and exit as well as aid in the production of RNA. Additionally, RNAPII interacts with other transcription regulatory proteins during initiation, elongation and termination for proper RNA formation and processing.

2. RNAPII Initiation

A great deal of work has been focused on RNA Polymerase II (RNAPII) transcription initiation which has led to the discovery of many proteins involved in transcription that stabilize and localize RNAPII at promoter regions to help in the initiation of RNAPII at protein coding genes specifically at transcription start sites (TSS) [18]. During initiation, general transcription factors TFIIB (Sua7), TFIID, TFIIE, TFIIF, TFIIH are recruited to the DNA promoter to form a pre-initiation complex (PIC). The initial step of forming the classical PIC is the binding of the TATA-box binding protein (TBP) to the consensus sequence TATAWAWR ~30 bp upstream of the TSS, which bends the DNA to prepare it for initiation [19]. A large percentage of promoters do not

contain TATA box sequences 80-90%; however, TBP is still present in these regions along with 13-14 TAF proteins, which help in identifying other promoter sequences, and together form the TFIID complex [20-22]. TFIID is conserved from yeast to humans and is involved in the recognition of core promoter sequences. TFIIA consists of proteins Toa1 and Toa2 and although not required for basal transcription it helps to stabilize the TFIID/DNA complex [23]. TFIIB (Sua7) is required for the recruitment of RNAPII through TFIIBs N-terminal domain and also helps stabilize the TFIID-DNA complex, through the TFIIDs C-terminal domain, and helps the promoter DNA open for the loading of RNAPII [23-26]. TFIIF (Tfg1/2/Taf14) is thought to stabilize and add specificity of RNAPII to its interaction with DNA [27]. TFIIE (Tfa1/2) binds to RNAPII and is required for the DNA to open at the promoter and also facilitates the recruitment of TFIIH [28-30]. The TFIIH complex is composed of a 7-subunit core, with helicase activity, and a 3-subunit protein kinase module, which includes kin28 and is responsible for phosphorylating serine 5 and serine 7 of the RNAPII CTD [28, 31-36]. This complex is essential for *in vivo* initiation and transcription and promotes processive transcription in a variety of ways including its DNA dependent ATPase activity, promoter opening and escape, and nucleotide excision and repair [37-41].

In addition to the primary pre-initiation complex (PIC) components, there are several other protein complexes and mechanisms that impact initiation and sometimes are involved in elongation and/or termination. The first of these is the nucleosome that is composed of a core of four proteins called H2A, H2B, H3 with H4 and 147 nucleotides of DNA wrapped around the core, which acts as a roadblock for transcription to occur [42-44]. In order for the nucleosomes to be assembled and disassembled at the

appropriate locations, several proteins have to interact with the histones. These include chromatin remodelers, histone chaperones, and histone variant readers which sometimes interact with histone tails, often removing or adding post translational modifications (PTMs) to induce a conformational change leaving regions of DNA open for other transcription factor proteins to bind to the DNA in a nucleosome free regions (NFR) [45-47]. NFRs are often found at promoters and enhancer regions of genes and are necessary for consistent RNA production from those particular genes [48]. These NFRs often allow several different transcription factor proteins to bind to the same region of DNA. These proteins can then also interact with other regulatory proteins that help activate transcription. One such complex of regulatory proteins called the Mediator is thought to bind to and activate several DNA sequence-specific transcription factors (TFs), possibly at the same time to facilitate initiation.

The Mediator complex is made up of a core of 26 proteins and the four subunit cyclin dependent kinase module called cyclin dependent kinase 8 (CDK8) module (CDK8, CCNC, MED12, MED13) [49]. Mediator is necessary for the formation of the PIC through the many contacts it has with RNAPII and helps recruit TFIIH, but is also involved in mediating several other TFs [50]. The catalytic subunit of TFIIH (Kin28) phosphorylates serine 5 of the CTD and results in the eviction of Mediator from RNAPII and thus initiation of transcription [51, 52]. Mediator is also directly involved with RNAPII and has been shown to bind to the CTD, but without the CDK8 module; mediator is thought to have two conformations, a core only mediator complex and a CDK8 module plus core complex. The CDK8 module together with core Mediator helps recruit the super elongation complex (SEC) which phosphorylates serine 2 on the CTD

via the CDK9 (Bur1) kinase [53-59]. These finding clearly established that the Mediator can act on transcription at initiation as well as during elongation.

3. RNAPII Elongation

Transcription elongation is a highly regulated phase of the transcription cycle that has become better understood in the last decade. In metazoans, transcription has been shown to pause at a significant number of RNAPII target genes soon after initiation at 20-120 nucleotides downstream of the TSS [60]. Pausing has been shown to occur through the action of the two complexes DSIF and NELF [61]. NELF is a complex only found in metazoans and is made up of 4 subunits: the core NELF-A and NELF-C which associates with NELF-B and NELF-D [62]. All four subunits of NELF have the capability to interact with RNA which could be part of a possible mechanism for pausing [62, 63]. NELF also appears to stabilize a paused RNAPII and inhibits premature termination [64, 65]. DSIF is composed of two proteins Spt4 and Spt5, which are conserved from yeast to humans. There is an Spt5 homolog in bacteria NusG [66-68]. Spt5 has a CTD domain which can be phosphorylated by TFIIH, as well as P-TEFb, and likely pauses RNAPII by interactions with nascent RNA at the exit channel of RNAPII [69-71]. Spt5 is also involved in the recruitment of the capping complex together with the phosphorylation of serine 5 on the CTD [72].

Messenger RNA processing mostly occurs co-transcriptionally, and begins as soon as the 5’ end of the transcript emerges from the RNAPII exit channel (Figure 2) [73]. The 7-methyl guanosine cap is added, which is important for stability of the RNA and also helps during RNA export [73-75].

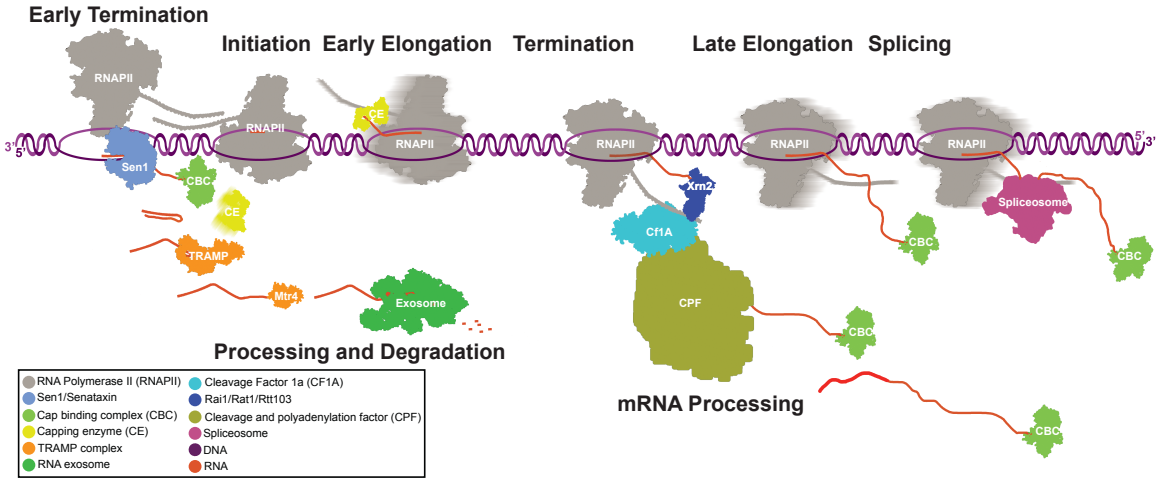


Figure 2: Overview of the different stages of transcription. A simplistic preview of the different stages of transcription are depicted. Blurry RNAPII complexes are meant to simulate an actively elongating process. Other RNAPII complexes depict a paused or stopped polymerase[76].

Capping is carried out by three enzymes in yeast, Cet1 which hydrolyzes the gamma phosphate of 5' RNA triphosphate, Ceg1 which adds guanosine monophosphate to the RNA diphosphate end and Abd1 which methylates the terminal guanine [77-80].

In mammalian cells, capping enzyme (gene name RNGTT) is bifunctional with both the 5' triphosphatase activity and mRNA guanylyltransferase activity encoded in a single protein [81-83]). In addition, a cap binding complex (CBC) which consists of Cbc2 and Sto1 in yeast, binds to the mRNA cap to further stabilize the RNA and perhaps provides feedback to stimulate transcription elongation [84, 85]. Co-transcriptional processing can also serve to signal for the decay of the nascent RNA such as when mRNA is not properly capped [86, 87]. Splicing can also occur co-transcriptionally, though not exclusively. It is possible that most spliceosomes (a large complex composed of approximately 80 proteins that is responsible for splicing introns from the premature mRNA), assemble on the nascent RNA transcript [88, 89].

Another interesting elongation complex, polymerase-associated factor 1 (Paf1C) (Cdc73, Ctr9, Leo1, Paf1 and Rtf1), is involved in multiple stages of transcription. Paf1C inhibits release of a paused RNAPII, but once released Paf1C positively regulates elongation of RNAPII [90, 91]. In addition, the Rtf1 subunit of Paf1C promotes the monoubiquitylation of H2B which influences chromatin structure during transcription [92]. Finally, to release the paused RNAPII, the positive transcription elongation factor complex (P-TEFb) phosphorylates serine 2 of the CTD, the CTD of Spt5, and the NELF complex [60, 93]. P-TEFb is a kinase complex consisting of CDK9 (Bur1) and cyclin T, and is a constituent of 3 bigger complexes including the SEC complex and is better known for the kinase activity of subunit Bur1. Once phosphorylated, NELF dissociates

from chromatin while a phosphorylated DSIF turns into a positive regulator of transcription elongation [94-96]. To continue elongation, nucleosomes must be removed from the chromatin downstream of RNAPII, and one of the complexes involved is the facilitates chromatin transcription complex (FACT). FACT (Spt16/Pob3) interacts directly with the H2A-H2B dimer to disassemble it from the DNA, allowing RNAPII to continue elongation [97, 98]. Other nucleosome interacting proteins/complexes such as Spt6, Chd1, COMPASS complex, Set2D, Bre1-Rad6, Dot1, and ISWI complex also take part in modifying the nucleosome for the progression of RNAPII [99-104].

4. RNAPII Termination

While initiation and elongation are critical, termination is likewise important for the proper RNA species to be stabilized and translated into protein. Variations in the end final stages of RNA production could cause major alterations in the mRNA 3' end stability and/or the turnover and recycling of RNAPII and therefore be disastrous for the cell. In order for transcription termination to occur at the appropriate location the RNA must be cleaved at the appropriate location, and then must be properly polyadenylated to avoid inappropriate mRNA degradation either within or outside of the nucleus of the cell and, in the case of mRNA, eventually be translated to the correct protein sequence. Currently there are well-described working models of eukaryotic transcription termination. One of these is the poly-A-dependent termination pathway which relies on the cleavage and polyadenylation factor complex and the cleavage factor complex (CPF-CF) termination factor. The CF complex consist of two sub-complexes CFIA (Pcf11, Rna14, Rna15, and Clp1) and CFIB (Hrp1). Hrp1, also known as Nab4, is required for

pre-mRNA cleavage and binding upstream of the UAUAUA sequence [105]. The CFIA complex is required for mRNA cleavage but also has a role in interacting with serine 2 of the CTD through the protein Pcf11 [106-109]. Additionally, CFIA may have a role in 3' end processing of small nucleolar RNA, snoRNA, which is required for guiding chemical modifications of other RNAs [106-109]. The CPF complex consists of 15 proteins Cft2, Mpe1, Ysh1, Glc7, Pta1, Pti1, Ref2, Ssu72, Swd2, Syc1 Cft1, Fip1, Pap1, Pfs2, and Yth1. The CPF complex contains the two major enzymes that perform the cleavage, Ysh1, and polyadenylation, Pap1. The other subunits likely have a variety of mechanistic roles but are also known to help stabilize and localize the position of the RNA of pre-mRNA cleavage [110, 111]. *In vitro*, the three budding yeast complexes CFIA, CFIB and CPF are sufficient to complete the process of cleavage and polyadenylation of pre-mRNA in budding yeast [112, 113]. Following polyadenylation, the mRNA is bound by poly(A) binding protein Pab1 and Nab2 which together stabilize the mRNA and guide its export out of the nucleus along with the cap binding complex described earlier [114, 115].

Transcription termination also requires that RNAPII be removed from the single stranded DNA template for which two models of the mechanism of action have been proposed. The first is the allosteric model that suggests that there is a conformational change that occurs after elongation factors dissociate from RNAPII to facilitate template release. The other is the "torpedo model" which involves a more direct approach. Rat1, a 5'-3'exonuclease, would degrade the remaining nascent RNA that remains on the RNAPII dislodging it from the template DNA strand with the help of its interacting partner Rai1, a decapping endonuclease, and a CTD interacting protein Rtt103 [116-120]. While Rat1 and Rai1 are involved in the termination mechanism, they are also involved

in the general degradation of unstable RNA [121-124]. Additionally, there is a distinct RNA quality control mechanism that is active in the cytoplasm and also involves an RNA decapping complex (Dcp1 and Dcp2) as well as another 5'-3' exonuclease Xrn1 [125-127]. There is also a major complex that degrades improperly processed pre-mRNA or unstable RNA in the nucleus and in the cytoplasm called the exosome [128-132]. The exosome is composed of a core of nine proteins in a stacked ring conformation. The bottom ring of the core contains Rrp41, Rrp42, Rrp43, Rrp45, Rrp46 and Mtr3 while the top ring has Rrp4, Rrp40 and Csl4 [133]. The exosome curiously has two 3'-5' exonuclease enzymes Dis3 and a nucleus specific Rrp6 [134].

Another termination mechanism is called the poly-A-independent termination pathway or the NNS termination pathway, which is so named after three proteins, Nrd1, Nab3 and Sen1. Nrd1 has a CTD interaction domain which binds to a phosphorylated serine 5 CTD during early elongation and has an RNA recognition motif as well as a Nab3 interaction domain [135-137]. Nab3 also has an RNA recognition motif and an Nrd1 interacting domain and forms a heterodimer with Nrd1 [138-140]. Sen1 is an ATP dependent RNA-DNA helicase that is thought to stimulate termination by disturbing the transcription elongation complex [137, 141-146]. This pathway is thought to mainly affect ncRNA but appears also to affect mRNA [147-149]. It has been suggested that NNS termination is not RNA-type dependent but length-of-transcript dependent, often occurring around 90-450 nt after the start of transcription [150, 151]. NNS termination is coupled to the processing and degradation of RNA transcripts via Nrd1 and its recruitment of the Trf4/5, Air1/2 and Mtr4 polyadenylation complex (TRAMP) and the nuclear exosome [12, 152, 153]. Trf4/5 are poly-A polymerases which add a short

adenine tail at the 3' end of RNA after NNS termination. Mtr4 is a helicase and can also help in managing the size of the adenine tail [154]. The nuclear exosome then either trims the adenine tail or completely degrades the RNA depending on the type of transcript and other factors [155]

5. Pervasive Transcription

It is important to understand all steps of transcription including some that might not occur in the standard cycle of transcription. Recent work has shown that many regions of DNA previously thought to be transcription inactive are undergoing what has been called "pervasive transcription" [12, 156-158]. One study found evidence that 85% of the genome in yeast is being transcribed [156]. In addition, there is evidence that a major source of pervasive transcription occurs at bi-directional promoters and at NFRs, resulting in a plethora of ncRNAs such as: stable unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs) [12, 13, 159, 160]. The function or purpose of many of these emerging ncRNA classes in the nucleus has yet to be completely understood. Nonetheless it is known that the production of ncRNA transcripts is dependent on a number of different core components of the basal transcription machinery [34, 135, 137, 161-164].

Transcription Elongation and the C-terminal Domain

The process of RNAPII transcription occurs through different mechanisms, that are known to be coupled to RNA processing, which are regulated through the recruitment of proteins to the C-terminal domain (CTD) of the largest subunit of RNAPII, Rpb1[34, 137, 149, 161-164].

1. The Structure of the C-Terminal Domain of Rpb1

Rpb1 subunit of RNAPII contains the CTD, which is made up of recurring heptad amino acid repeats, YSPTSPS, which, when phosphorylated in various patterns, recruits different transcriptional proteins (Figure 1) [34, 135, 137, 161-165]. Of the seven amino acids, five residues can be phosphorylated and the phosphorylated forms will be referred to by as follows: tyrosine-1 (Y1P) serine-2 (S2P) threonine-4 (T4P) serine-5 (S5P) and serine-7 (S7P). At any time during transcription, different patterns of amino acid phosphorylation likely occur, but the precise combinations at any given point in transcription across all heptad repeats is not known. Of the seven residues in each heptad, serine 5 and serine 2 were found to be the most extensively phosphorylated in mass spectrometry experiments and *in vitro* studies [166-169]. In addition, the two prolines can undergo cis-trans isomerization. The modifications of all seven amino acids gives the CTD a variety of distinct surfaces for different proteins to bind [170]. The CTD is conserved across eukaryotes, but different organisms contain a different number of the repeating heptads. Humans and most vertebrates have 52 repeats, while *Drosophila* has 44, *Arabidopsis* has 40 repeats and *Saccharomyces cerevisiae* 26 heptad repeats. The CTD is essential and in *S. cerevisiae* at least 8 heptad repeats are required for viability and any Y1P/S2P/S5P substitutions are lethal [163]. There is amino acid sequence variation among the repeats with the majority of the variation occurring on the 7^{th} amino acid possibly allowing for the different heptads to have regional specificity to certain proteins [170]. Whether or not only one heptad is necessary to for all CTD interacting proteins is still up to debated. However, one study discovered that any addition of an

alanine residue between any heptad in the CTD was lethal in yeast, but up to five additional alanine residues could be added in between 2 heptads and still be viable [171].

As noted, isomerization of prolines 3 and 6 can also affect the structure state of the CTD. Proline isomerization is carried out by peptidyl-prolyl cis/trans-isomerases (PPIases) [172]. In yeast, this isomerization is carried out by Ess1 which recognizes the motif of a phosphorylated Ser5-Pro6 [173]. The proline-6 then goes from a Cis conformation to a trans conformation that makes the CTD more energetically favorable for certain proteins to bind to therefore adding one more dimension of complexity and specificity to the CTD [174, 175].

The CTD appears to have a residual secondary structure and it has β-turn composition when bound to other proteins, but it is generally flexible and contains a linker region that connects the heptads to the rest of Rpb1 [176, 177]. Previous studies have focused on the incomplete CTD by using short heptad repeats, and there is still a lack of information of the complete CTD's true structure in vivo. In these shortened CTDs, phosphorylation gives the CTD additional hydrogen bonds that appear to contribute to β-turn structures in a protein bound CTD. In full length CTD Sucrose gradient ultracentrifugation studies phosphorylation suggested an extended CTD conformation which may be essential for CTD protein interactions [178]. More recently, another group has found expansion of the overall volume of the CTD after phosphorylation but also found a reduction in flexibility as well as structural heterogeneity throughout the CTD using Small Angle X-ray experiments (SAXS) experiments [179]. These data provide further evidence that different sections of the CTD may be influencing transcription in different ways.

2. Phosphorylation Patterns of the C-Terminal Domain of Rpb1

Phosphorylation patterns of the CTD are indicators of the relative stage of RNAPII during transcription from initiation to termination [34, 162, 164, 180, 181]. Although the CTD is required for viability in cells, the CTD is not necessary for the catalytic activity of RNAPII [182]. This provides evidence that the CTD is not necessary for the synthesis of RNA but for the regulation of the process of transcription. By using phospho site specific antibodies, a number of research groups have mapped the relative abundance of phosphorylation of each phosphorylated residue group across the transcription cycle using chromatin immunoprecipitation based approaches (Figure 3) [35, 164, 180, 183-185]. Tyrosine 1 is essential for the viability of a *S. cerevisiae* cell as the substitution of Tyrosine 1 with a phenylalanine is lethal [163].

Generally, Tyrosine 1 is phosphorylated during early elongation by c-Abl in humans and then is dephosphorylated by Glc7 near the termination site; phosphorylation is present on RNAPII located at most active genes [151, 186, 187]. Fluorescence anisotropy experiments by Mayer et al. showed that Tyrosine 1 phosphorylation leads to a decreased ability of CTD peptides to bind to recombinant proteins Pcf11, Nrd1 and Rtt103, the only known proteins with CTD interaction domains (CID), which are involved in different mechanisms of termination [151]. Y1P; however, appears to increase CTD peptide binding to Spt6, an elongation protein that binds a S2P CTD through its Src homology domain (SH2) [188].

Serine 2 is phosphorylated in *S. cerevisiae* by Bur1 and Ctk1 during early elongation and gets dephosphorylated by the Fcp1 phosphatase before the next round of transcription [189, 190]. S2P is essential and aligns with periods of elongation, splicing,

termination, mRNA processing and export [118, 174, 191-197]. There are two known termination factor proteins, Pcf11 and Rtt103, that have a CID specific for a S2P CTD [118, 174].

Threonine 4 is phosphorylated during mid-stage elongation and is dephosphorylated right before the polyadenylation site is reached [151]. Currently there is no known threonine 4 kinase in yeast, but in human cells Plk3 has been shown to phosphorylate threonine 4 [198]. While T4P is not essential for viability in yeast, it is essential in chicken and human cells [171, 199, 200]. The mutation of threonine to valine leads to decreased processing of non-polyadenylated histone mRNA as well as defects in elongation [198, 200].

Serine 5 of the CTD heptad is essential and is the most phosphorylated residue. Serine 5 is phosphorylated early in transcription and is dephosphorylated as RNAPII nears termination [151, 166, 181]. Serine 5 is phosphorylated by Kin28 a subunit of TFIIH in *S. cerevisiae,* and dephosphorylated by both Rtr1 and Ssu72 [34, 201-207]. Serine 5 phosphorylation is associated with initiation, the dissociation of the mediator complex, promoter release, mRNA capping and early transcription termination via Nrd1 [83, 135, 136, 208-211].

Serine 7 begins to be phosphorylated at the beginning transcription and phosphorylation levels stay relatively constant throughout the gene in transcription process [183, 212-214]. The removal of S7P is lethal in human cells but is not essential in yeast and a serine 7 to alanine mutation had no effect on mRNA but affects the correct processing of snRNA [33, 199, 215]. Serine 7 is phosphorylated by Kin28 during the initiation stage and chemical inhibition studies show that serine 7 stays phosphorylated

via Bur1 during the later stages of transcription [34-36, 204, 212]. Ssu72 phosphatase removes the serine 7 phosphate immediately after cleavage and polyadenylation [13, 184, 185, 201, 216, 217].

3. Transcription Elongation and the CTD

Like all stages of transcription, elongation is heavily regulated by the state of the CTD. RNAPII release from the promoter is facilitated by phosphorylation of Serine-5 of the CTD by Kin28 thus allowing RNAPII to transcribe through the gene. However, various other proteins are necessary to maintain RNAPII activity. Often RNAPII stalls or pauses during its activity, and phosphorylating the CTD can help RNAPII to begin to transcribe again. The structure of a paused mammalian RNAPII elongating complex shows that DSIF forms a clamp around upstream DNA and nascent RNA while NELF binds the polymerase funnel and opens the trigger loop that inhibits the binding of the next nucleotide triphosphate (NTP) [39, 61, 104, 218-220].

P-TEFb contains a CTD kinase Bur1 which phosphorylates Ser-2 of the CTD and also phosphorylates Spt5, NELF, Paf1C, Spt6 and the CTD linker which leads to the resumption of RNAPII transcription [94-96, 221, 222]. A recent cryo-EM study showed that P-TEFb activity facilitates the formation of a stable complex between RNAPII, DSIF, Paf1C, and Spt6 in vitro which may explain the mechanism of unpausing RNAPII [104]. Additionally, Vos et al. show that the interaction between Paf1C and Spt6 may induce a conformational change releasing the DSIF clamp and promote the rewinding of DNA at the transcription bubble promoting elongation [104, 223].

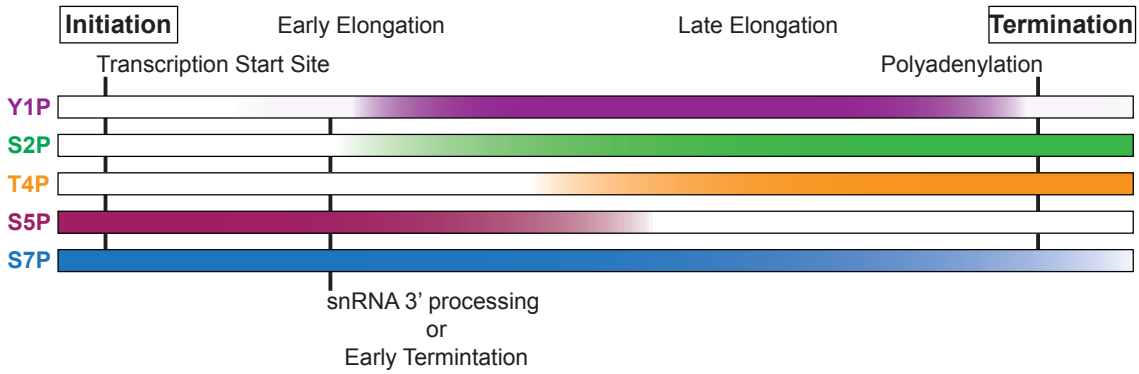


Figure 3: Phosphorylation pattern throughout transcription. Phosphorylation for each individual acid is mapped based on prevalence. The darker the color, the greater the degree of phosphorylation during the specified stage. Adapted from Heidemann, et al. [181]

This is consistent with previous studies showing that Spt6, which binds to S2P of the CTD, stimulates elongation in vivo and in vitro [224-227]. In addition, the tandem SH2 domain of Spt6 appears to bind to a doubly phosphorylated region in the linker of the CTD at a phosphorylated Serine (S1493) and at either threonine (T1471) or tyrosine (Y1473), which may be phosphorylated by P-TEFb (Ctk1 and or Bur1), that helps maintain repressive chromatin thus promoting elongation [104, 228].

Paf1C is also implicated in pause release and elongation through its interaction with Spt4/5 and the recruitment and activation of histone modifier FACT complex [90, 91, 224, 229]. The Cdc73 subunit of Paf1C interacts directly with a diphosphorylated (S2/S5 and S5/S7) RNAPII CTD through its Cdc73's own C-terminal domain *in vivo* and helps recruit Paf1C to RNAPII [230-233]. In addition to Rtf1's role in chromatin structure, it also interacts with Spt5, and phosphorylation of the CTD as well as Spt5's C-terminal repeats (CTRs) phosphorylation are required for proper recruitment of Paf1C [234, 235]. Another study showed that there is a mutual dependence on P-TEFb and Paf1C for their own recruitment to active chromatin in a positive feedback mechanism that promotes RNAPII pause release by regulating the association of P-TEFb on chromatin directly as well as through histone modifications via Paf1C [229, 236].

Transcription Termination

Transcription termination is an important regulator of the expression of non-coding and coding RNA transcripts [142, 148, 237-242]. Depending on the pathway, different RNA transcripts can be degraded, stabilized, or exported to the cytoplasm for translation [147, 148, 237]. Currently, there are two model types of RNAPII termination: poly-A dependent termination and NNS termination have been proposed (Figure 4).

1. Poly-A Independent Termination (NNS Termination)

Nrd1 has been shown to interact with the nuclear exosome enhancing the ability of the exosome, possibly by recruiting the exosome to the site of termination, to degrade RNA transcripts *in vivo* [155]. Interestingly the majority of small ncRNAs are terminated during early elongation when serine 5 phosphorylation is most abundant on the CTD, [135].

Nrd1 also interacts directly with the TRAMP complex, which polyadenylates RNA transcripts, through poly(A) polymerase proteins Trf4/5 [12, 152, 155]. The TRAMP enhances the ability of the exosome to degrade aberrant RNA *in vitro* [152]. Further studies have shown that the TRAMP complex increases the activity of Rrp6 independent of the exosome [243]. These studies provide evidence that the NNS pathway is involved in termination, trimming and degradation of short non-coding RNA transcripts with unstable RNA usually being completely degraded. A faulty exosome or deficient NNS termination can lead to over-abundance of ncRNA or RNAPII termination failure that creates issues downstream such as the suppression of nearby genes, also known as transcription interference ([237, 244-249]. Most small ncRNAs, including sno/snRNAs, are terminated via the NNS termination pathway in *S. cerevisiae* which generally leads to non-coding transcripts being either degraded or processed [12, 130, 137, 147, 148, 159, 165, 237, 244, 250]. A number of additional protein complexes regulate this process including the TRAMP (Trf4/5– Air2/1– Mtr4 polyadenylation) complex and the nuclear exosome that work together to process and polyadenylate small functional ncRNAs, such as snRNA and snoRNA [138, 142, 147, 155, 165, 251, 252].

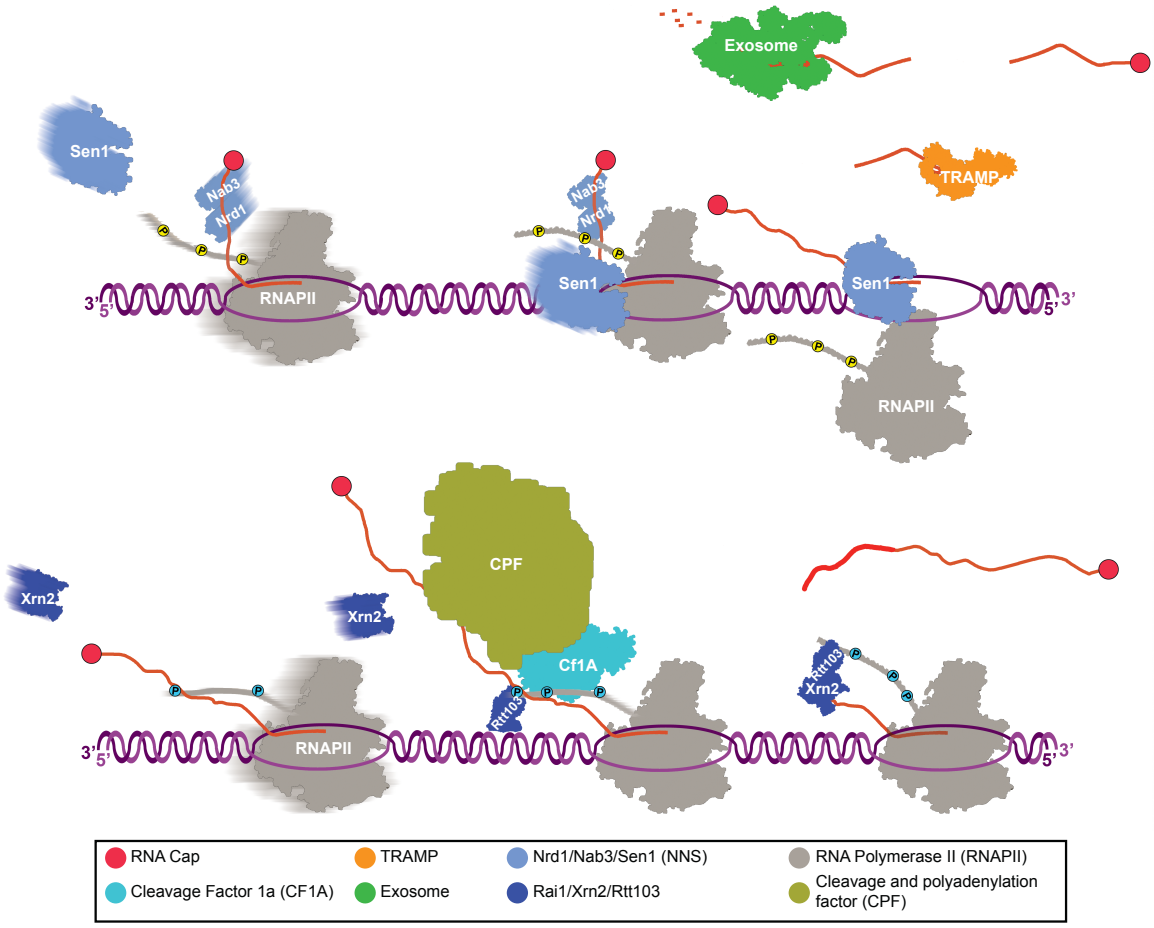


Figure 4: Models of RNA Cleavage-Independent/Dependent Transcription Termination. (A) In yeast, the RNA cleavage-independent pathway is also known as the NNS pathway. Nrd1 binds to a serine 5 phosphorylated CTD and Nrd1 and Nab3 bind to RNA. Sen1 is a DNA/RNA helicase. The TRAMP complex polyadenylates the 3'end of RNA, and the exosome (green) processes or degrades RNA. (B) In the Poly-A dependent pathway, CF1A is recruited to Ser2 phosphorylated CTD. CPF is recruited and cleaves RNA at the polyadenylation signal. Poly(A) polymerase polyadenylates 3' end of RNA following cleavage. Xrn2 degrades the 5' end of uncapped RNA stimulating RNAPII removal from the DNA template.

The exosome is responsible for degrading nonfunctional small ncRNA such as CUTs, which generally terminate at about 450 nucleotides after the transcription start site (Table 1, [12, 148, 150]). SUTs are stable RNA transcripts which are transcribed from regions of the genome which were previously unannotated (Table 1, [159]). CUTs were first detected in microarray experiments after inhibiting the NNS-coupled degradation pathway in a *S. cerevisiae* strain through the deletion of one of the nuclear exosome 3'-5' exonuclease subunits, *RRP6* [12].

Gene expression is regulated by early termination at specific RNAPII target genes, which has been termed transcription attenuation [142, 237-242, 246, 253, 254]. Transcription attenuation of mRNA often occurs at genes which exhibit antisense transcription at protein coding genes [237, 246, 254]. Antisense transcripts (ASTs) are transcripts of RNA that are produced from the non-coding strand of a protein coding gene. Another study discovered that while 80% of *S. cerevisiae* open reading frames (ORFs) had an RNA transcript antisense to the ORF, 30% of ORFs in human liver cells had RNA antisense to the ORF [255]. Several studies have classified different categories of transcripts and have described their mechanisms of termination and degradation (outlined in Table 1). Altering the function of proteins involved in NNS termination has led to the discovery of hundreds of ncRNAs, resulting in global changes in the transcriptome [12, 13, 130, 159, 160]. Termination is less recognized in the field of transcription as a form of regulation of RNA expression than other stages of transcription, such as initiation; however, these and other studies suggest that the termination machinery plays a major regulatory role in the control of gene expression.

2. Poly-A Dependent Termination (CPF-CF Termination)

The polyadenylation dependent transcription termination pathway is dependent on the CPF and the CFIA/B complexes for the termination and processing of mRNA [112, 117, 118, 237, 256-260]. Few details of the exact mechanism of transcription termination are known is still unknown. Many proteins involved recognize termination signals and cross-talk through mechanisms such as CTD phosphorylation. For example, Glc7, of the CPF, is required for the dephosphorylation of tyrosine 1 of the CTD near the AAUAA poly(A) termination site. This may help recruit the CF1A subunit Pcf11, and Rtt103 since it is thought that tyrosine 1 phosphorylation inhibits the ability of these proteins to bind to serine 2 [151, 261]. Pcf11 contains a CTD interaction domain (CID) that is specific to a serine 2 phosphorylated CTD at the 3' end of protein coding genes during transcription termination [109, 118, 174, 180, 181, 183]. Rtt103 contains a CID domain that is also specific to serine 2 phosphorylated CTD and is thought to help recruit CFIA factors along with Pcf11 [117, 118]. After recruitment, CFIA proteins RNA14 in complex with RNA15 bind the G/U rich region of RNA near the site of termination [262]. In addition, RNA14 has been shown to specifically interact with the CPF protein Cft1 [263]. CFIA and CPF complex appear to operate in a concerted manner where by Fip1 binds to U-rich regions near the site of termination, and Cft1 and Yth1 bind to the termination site specifically at the polyadenylation poly(A) site, AAUAA, to induce pausing of RNAPII [263-267]. During the pause another CPF protein the endonuclease, Ysh1, which cleaves the 3' end of the newly transcribed RNA at the poly(A) site [110, 268, 269]. The newly cleaved RNA then becomes polyadenylated at the 3' end by the Pap1 enzyme, and requires CPF proteins Cft1, Cft2, Ysh1 and Yth1 for poly(A) site specificity [111, 270, 271].

Transcript	Termination Pathway	Stability	Degradation Factors
mRNA	Poly-A Dependent	Stable	None
snRNA/snoRNA	NNS(Nab3-Nrd1-Sen1)	Stable (3' end processed)	TRAMP, Rrp6, Exosome
CUT~440 nt	NNS(Nab3-Nrd1-Sen1)	Unstable	TRAMP, Rrp6, Exosome
SUT ~760 nt	Poly-A Dependent/ NNS(Nab3-Nrd1-Sen1)	Stable/Unstable	TRAMP, Rrp6, Exosome
SRT	NNS(Nab3-Nrd1-Sen1)	Unstable	Unknown

Table 1: Transcript classes and characteristics. Table includes names of transcripts with the mechanism in which they are terminated and degraded as well as their stability in a wild-type cell. RNA classes are cryptic unstable transcripts (CUTs). stable unannotated transcripts (SUTs) and Ssu72 restricted transcripts (SRTs). *Table adapted from: Odil Porrua and Domenico Libri, Nature Reviews Molecular Cell Biology* 16, 190–202 (2015) [12, 13, 159, 181, 240, 272-274]

CPF protein Fip1 also has been shown to interact with and stimulate Pap1 activity [267].

Although there has been much progress in understanding the mechanisms of termination, but there is still much that is not yet known. The CPF complex contains 15 protein members, but many of their exact functions are not yet known even despite evidence that all are involved in the 3' end processing of the nascent protein coding RNA or termination [275]. Many of the CPF proteins form distinct subcomplexes within the complex such as the associated with Pta1 (APT) complex which consists of Swd2, Ref2, Glc7, Pti1, Ssu72, Pta1, and Syc1. This assembly was discovered through affinity purification experiments [256]. The APT complex contains many proteins whose functions in 3' end processing of RNA is not understood. However, proteins like the phosphatase Ssu72 are known to be involved in other important transcription processes such as serine 5 and serine 7 CTD dephosphorylation, gene looping, and read-through of termination of ncRNA [13, 185, 201, 276, 277]. Swd2 is also known to be a subunit in the Set1C histone H3K4 methyltransferase complex, and although it does not participate in polyadenylation of the RNA, it is thought to cause defects in termination [278, 279]. Ref2 is an unspecific RNA binding protein that appears to increase stability of the protein coding 3' end processing proteins at inefficient poly(A) sites as well as at ncRNA 3' end sites [280, 281]. Pti1 also was found to be involved in the 3' end processing of ncRNA [281]. Pta1 is required for cleavage and polyadenylation of the protein coding RNA and is thought to facilitate the assembly of other CPF subunits that are directly involved in the cleavage and polyadenylation of RNA [112]. Syc1, which is homologous for the C-terminus of Ysh1, was recently shown to be necessary to create an exclusive sub complex of CPF that contains only the APT members of the CPF; removal of Syc1 increases

cleavage and polyadenylation activity [269, 282]. Syc1 also appears to preferentially crosslink to small nuclear RNA (SNR) genes, and removal of Syc1 reduces SNR transcript expression in the cell [282]. Thus, all these data are suggestive that the CPF may work in two or more separate subcomplexes that include different sets of proteins to process different classes of transcripts. A lot of key information is missing to fully understand the mechanism of either poly(A)-dependent or poly(A)-independent termination. Recently a technically advanced method has been able to separate and identify sub-groups of the CPF complex and more accurately identify which proteins are in these different groups. Using cryogenic electron microscopy (cryo-EM) and non-covalent nanoelectrospray ionization mass spectrometry (nanoESI-MS), the Passmore lab was able to separate the CPF complex into 3 modules named APT/phosphatase(Ssu72, Pta1, Syc1, Swd2, Glc7, Ref2, Pti1), nuclease(Ysh1, Cft2, and Mpe1), and the poly(A) polymerase module(Pap1, Cft1, Yth1, Fip1, and Pfs2) [283, 284].

CTD Phosphatases

Known regulators of the CTD include a number of phosphatases and kinases. Some have multiple targets while others have only one known target. The known serine kinases in S. cerevisiae include Kin28(S5P/S7P), Bur1(**S2P**/S5P/S7P), and Ctk1(**S2P**/S5P/S7P). The known serine phosphatases include TFIIF associating component of CTD phosphatase, Fcp1(**S2P**/S5P), regulator of transcription 1, Rtr1 (S5P), and suppressor of Sua7 gene 2, Ssu72(**S5P**/S7P). Like the kinases the phosphatases play an equally important role in regulating the CTD and consequently RNA transcription. S5P is essential residue of the CTD, and serine 5 is the most highly phosphorylated

residue [166, 168]. There is substantial evidence that the phosphorylation state of serine 5 of the CTD plays a significant role in the fate of transcription, especially with short ncRNA transcripts, which are regulated by the NNS pathway; yet the understanding of the major players of transcription and their specific roles are still unclear.

1. Fcp1 Phosphatase

Fcp1 is an essential protein in yeast and was the first CTD phosphatase to be identified to convert a phosphorylated form of RNAPII to an unphosphorylated version [285-287]. Fcp1 can dephosphorylate both S2P as well as S5P but displays a ~6-fold preference for S2P and prefers an S2P-P3 motif to be in trans conformation [173, 288, 289]. Mutation of Fcp1 in the CTD binding pocket that reduce specificity for S5P did not result in growth defects in yeast; however, reduction in S2P does have an effect on yeast growth insinuating that Fcp1's main purpose is to dephosphorylate S2P and its ability to dephosphorylate S5P may be nonspecific [189]. Fcp1 has also been shown to interact with the catalytic core of RNAPII as well as TFIIF, which is thought to stimulate its phosphatase activity [286, 290-293]. The presence of Fcp1 stimulates transcription *in vivo* and *in vitro* and work done in John Lis's group has shown that depletion of Fcp1 leads to an accumulation of nonchromatin bound phosphorylated RNAPII and decreasing amounts of unphosphorylated RNAPII, which is required to initiate transcription [189, 294, 295].

2. Rtr1 and Ssu72 Phosphatases

Currently, there are two main phosphatases that target phosphorylated serine 5 Rtr1 and Ssu72 [34, 201, 203, 205-207]. *SSU72* is as an essential gene conserved across eukaryotes, and *in vivo* studies have found that Ssu72 and the scaffold protein Pta1 are a part of the CPF complex [13, 216, 256, 276, 277, 296-300]. Several experiments have linked Ssu72 to separate mechanisms of transcription including: initiation, RNA 3' end processing, termination, cleavage and polyadenylation in complex with CPF, elongation, and gene-looping [13, 256, 276, 277, 297-300]. Ssu72's role in transcription has even led one group to describe a set of transcripts, which they labeled as Ssu72 restricted transcripts (SRTs), they discovered using tiling arrays in an Ssu72-R129A mutant which were not present in the cell in the wild-type [13]. Rtr1 and the human homolog RPAP2 were initially found to interact with RNAPII in mass spectrometry experiments; in addition, ChIP-seq assays revealed that Rtr1 localized at protein coding genes at similar regions that ChIP-seq experiments identified where serine 5 was phosphorylated [203, 301]. Rtr1 is a phosphatase that has been shown to dephosphorylate serine 5 *in vivo* and *in vitro* using serine 5 specific antibodies [203]. The phosphatase activity of Rtr1 was recently confirmed with a crystal structure that reveals a binding pocket with a trapped sulfate ion, a potential site for phosphate group binding [205]. Irani *et al.* identified an amino acid substitution, Y105A, in this binding pocket that reduced the phosphatase activity 40-fold [205]. Using ChIP-chip experiments with a serine 5 phosphorylated CTD antibody one group was able to quantify serine 5 phosphorylation at set regions along a gene based on the amount of DNA that was immunoprecipitated along with the S5P CTD antibody [302]. This group calculated the average number of DNA reads, which represent

the amount of S5P, found at each region of the genes in an *rtr1Δ* and WT strain. Their data displayed an increased average of S5P at the 3' end of the genes which implies that there is a global 3' shift in localization on average across all genes in the genome when Rtr1 is removed which gives further evidence that Rtr1 has a major impact in the distribution of serine 5 phosphorylation of the CTD [302]. Both phosphatases Rtr1 and Ssu72 appear to have the same serine 5 target, yet *S. cerevisiae* strains with an *RTR1* deletion or a Ssu72 mutation seem to have contrasting effects. ChIP assays using an Rtr1 antibody show that on most genes Rtr1 accumulates during early elongation while an Ssu72 antibody most often finds Ssu72 accumulating at the middle to late elongation [185, 203]. The difference in timing of phosphatase activity could be due to the necessity for proline 6 to be isomerized, by Ess1, into the cis state for Ssu72 to have a higher affinity for a serine 5 phosphorylated CTD and to be more efficient in the dephosphorylation of the serine 5 [184, 216, 217, 303-305]. However, since Rtr1 acts first when phosphorylated serine 5 is most abundant, Rtr1 might play a greater role in fine-tuning the activity of RNAPII, thus deciding the fate of most transcripts. Ssu72 might be more responsible for coarse tuning and RNAPII recycling which could be essential for transcription termination.

Mass Spectrometry

Today mass spectrometry is utilized for studying several fields of biology. In addition, the field of mass spectrometry is producing new methodology to answer questions that could have never been answered before as exemplified with multidimensional protein identification technology (MudPIT), affinity purification mass

spectrometry (AP-MS), and nanoESI-MS as mentioned previously in the CPF modules study [283, 284, 306].

1. MudPIT and AP-MS

The use of mass spectrometry in biological experiments started becoming more common after DNA and RNA sequencing became prevalent. Just as scientists began to look at mRNA differences in a cell, many started to realize that looking at the end product of mRNA could also be analyzed in a similar way. In addition, there is evidence that the abundance of mRNA and the corresponding proteins may not be correlated [307-310]. Thus, there was a push to be able to emulate RNA sequencing methodology with mass spectrometers. Some of the limiting factors that scientists sought out to address were; low abundance of certain proteins, the ability to detect membrane proteins, to detect proteins with very low/high isoelectric points, and to identify proteins with a very large or small mass [311-316]. Several methods were developed to overcome these problems including a more efficient separation of complex protein mixtures such as MudPIT([317]. By using this method, scientists can measure protein levels of whole proteomes and compare proteomes of different cells in different conditions or mutations as well as analyze more complex set of protein mixtures [318-323]. The MudPIT method utilizes a bi/triphasic microcapillary column packed with chromatographic phases, such as hydrophobic stationary or reversed-phase (RP) resin and strong cation exchange resin (SCX) that separate peptides based on their chemical and physical qualities. The column is then linked to a high-performance liquid chromatography (HPLC) instrument and coupled to a tandem mass spectrometer[317, 324]). The peptides are pressure loaded onto

the packed microcapillary column and then eluted into the mass spectrometer using an organic phase gradient such as acetonitrile with the HPLC.

Then the peptides are ionized by applying a continuous high voltage (2–6 kV) via a nano-electrospray ionization (nano-ESI) source to the eluting peptides [325-327]. The peptides are then fragmented in a predictable fashion once or multiple times in the tandem mass spectrometer via collision induced dissociation (CID), high energy collision induced dissociation (HCD) or electron transfer dissociation (ETD) [328, 329]. The resulting fragment ions can then be detected by their mass and charge using components inside the mass spectrometer such as an Ion trap, an Orbitrap, or ion cyclotron resonance (ICR) mass analyzer [330, 331]. The resulting data generates spectra for each peptide which is then analyzed by a computer, and each spectra profile is compared to a database of peptides of possible amino acid sequences in different species using an algorithm such as SEQUEST that produce peptide spectral matches (PSMs) and the probability that these PSM's are true matches [332].

AP-MS is another method that has been instrumental in many researchers' discoveries over the years [333-336]. This method allows scientists to analyze and map convoluted protein-protein interactions (PPIs) in a cell. Simplification of the proteome is achieved by isolating proteins of interest and then detecting the proteins that interact with them (Figure 5). The affinity purification can be altered to increase or decrease specificity that can serve to give a more complete picture of proteins that interact with the bait protein at the level of binding and those that have low transient interactions.

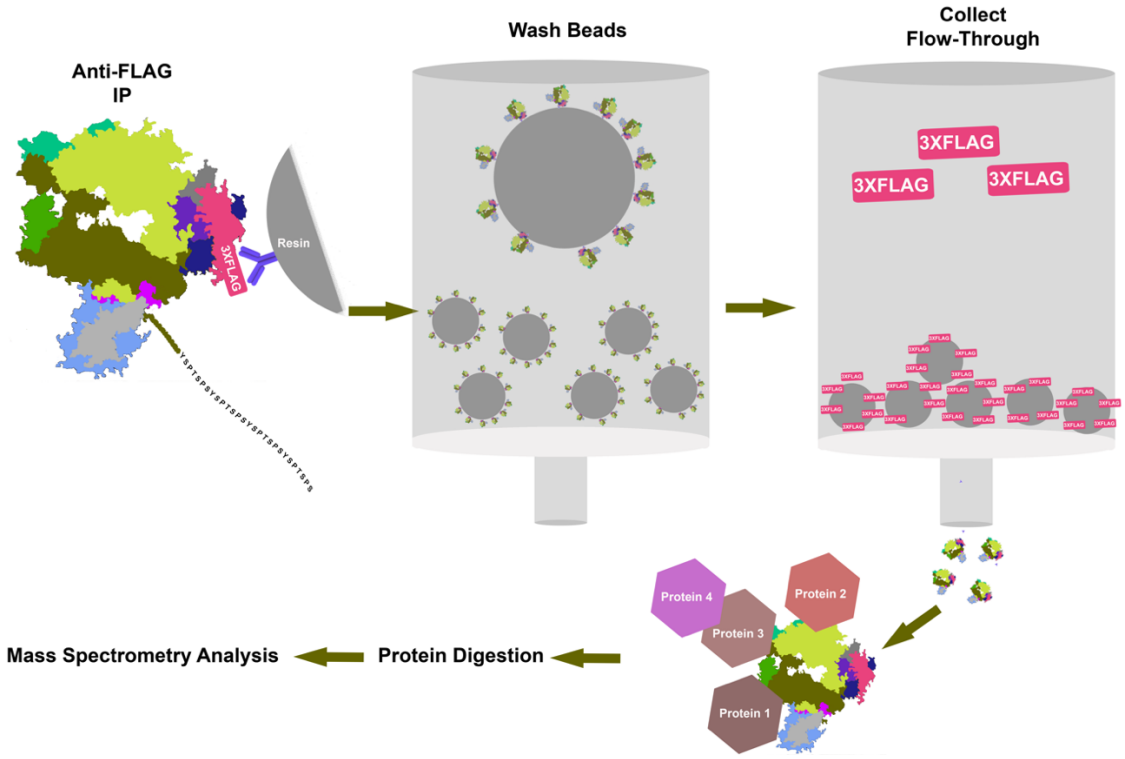


Figure 5: Affinity purification mass spectrometry schematic. Flag-tagged protein is immunoprecipitated using anti-flag beads, then washed on a column to remove non-specific binding of proteins. Flag peptide is added to remove immunoprecipitated proteins from beads through competitive elution. Proteins are collected and then digested using trypsin and Lys-c to create peptides for the mass spectrometer.

2. Cross-Linking Mass Spectrometry

Structural studies are incredibly useful at providing information about proteins; yet protein structures are difficult to solve, and they take a significant amount of time to complete, due to their difficulty in forming crystals, protein flexibility, low solubility and large size [337]. With increasing complexity the difficulty rises tremendously [337]. An emerging method to study the architecture of proteins and their interacting partners is to chemically cross-link the proteins of interest and then use mass spectrometry (XL-MS) to identify where proteins are cross-linking either to themselves or to other proteins [338-341]. In this method, proteins are chemically cross-linked, enzymatically digested and then the resulting sample is processed by liquid chromatography tandem mass spectrometry LC-MS/MS or LC-MS/MS/MS by ESI. Using this method, the mass spectrometer can identify which amino acids are cross-linked by identifying peptides which contain the mass of the crosslinker. The short length of the crosslinker gives this method the ability to identify where these proteins may be interacting thus giving insight to the structural architecture of the protein itself and in relation to other proteins. Identifications of interactions are restricted by the crosslinker and the amino acids that it can crosslink to. However, many different crosslinkers of varying specificity and length are currently available and many new crosslinkers are currently in development. This method is both fast and inexpensive and many different proteins and mutations of the same protein could be analyzed quickly. A limiting factor for XL-MS is the efficiency of the technique; thus, quantity of sample may be an issue. Some crosslinkers recommend to use between 10μM-500μM quantities of purified protein. For XL-MS, a commonly used crosslinker for many studies has been Bissulfosuccinimidyl suberate (BS3). BS3 requires

the computation of a pre-determined peptide library of possible crosslinks, which can lead to an incomplete/limited library of possible crosslinked peptides or a large computational task which would be unfeasible and/or inefficient for complex mixtures of peptides [342-345].

Another XL-MS method is one that uses a CID labile MS cleavable, cell membrane permeable crosslinker called Disuccinimidyl sulfoxide (DSSO), which has a spacer length of 10.1 Å [346, 347]. The DSSO crosslinker contains two amine reactive N-hydroxysuccinimide (NHS) ester groups at the ends of the molecule, which target lysine residues, and a sulfoxide in the spacer region which has the ability to be cleaved inside the mass spectrometer through low energy CID levels without cleaving the peptide bonds. NHS esters have also been observed to react with hydroxyl groups in serines, threonines, and tyrosines but only at 20% of their reactivity with lysines and so it may be possible to gain more information from these experiments where the number of crosslinks have been relatively scarce [348]. After cleavage each crosslinked parent peptide stays intact while leaving two distinct modifications on lysines, a carbonyl group with either a thiol group or an alkene group at the end. The difference in mass between the two groups is 32 daltons, this mass difference is used to identify true crosslinked peptides. The ability to cleave the crosslinker inside the MS and then identify crosslinked peptides as modifications of the original peptide give the crosslinker a huge advantage in that massive computing power or a computationally predetermined library of crosslinked peptides are unnecessary. The feature allows the DSSO crosslinker to be used to identify a more diverse and complex set of crosslinked samples.

Investigating the Role of Rtr1 and Ssu72 in RNAPII Transcription Termination

The highly regulated process of transcription requires an array of proteins and protein complexes. These proteins can be specific or non-specific, unique or redundant, essential or non-essential. In order to better understand the transcription process as a whole, it is important to understand all of the components of the process, and each component's individual role. The RNAPII CTD is an essential hub for recruitment of proteins and the regulation of RNAPII transcription. RNAPII CTD hyperphosphorylation at serine 5 is required for transcription to initiate, which then undergoes dephosphorylation as transcription progresses through the gene and is likely required for transcription termination to occur. Currently only two phosphatases, Rtr1 and Ssu72, are known to dephosphorylate serine 5 on the RNAPII CTD. However, their mechanism of action is still unknown and it has yet to be discerned whether these phosphatases target the same S5P CTD or whether their dephosphorylation incurs the same downstream effect on transcription.

Past research based on ChIP assays suggests that Rtr1 likely acts during early elongation, while Ssu72 acts in mid/late elongation [34, 201, 203, 296, 349]. ChIP assays with antibodies specific to Rtr1, serine 5 phosphorylation and Ssu72 show that Rtr1 and serine 5 phosphorylation are most often found to be associated with 5' end of genes, Ssu72 is predominantly associated with the 3' region [184, 185, 203, 217]. My central hypothesis is that Rtr1 and Ssu72 play a major and unique role in the regulation of early termination through dephosphorylation of the RNAPII CTD and thereby effect recruitment of the NNS termination pathway or other downstream proteins. In addition, I believe that Ssu72 plays an additional key role in transcription that is separate and downstream of Rtr1's role, specifically with Ssu72's associated complex CPF in PolyA

dependent termination**.** To study the effect of Ssu72 on transcription, I used a termination deficient strain of *Saccharomyces cerevisiae* originally named *ssu72-tov* (termination override) which I refer to as Ssu72-L84F [350, 351]. I studied Rtr1's role in transcription using an *rtr1Δ* yeast strain.

In order to study the direct effect that Rtr1 and Ssu72 have on the transcription process and the RNA product I used a termination defective Ssu72-L84F and *rtr1Δ* mutant strain to analyze differentially expressed genes with a particular focus on NNS termination targets. To analyze the phosphatases effect on RNAPII specifically and to confirm whether RNAPII continues to transcribe beyond different termination signals globally, in the Ssu72-L84F and *rtr1Δ* strain, I measured the abundance of RNAPII occupancy using a method called ChIP-exo, a high resolution chromatin IP (ChIP) method [352]. In order to study the protein network that Rtr1 and Ssu72 are involved in I used MudPIT-LC/MS/MS to analyze the effects that the Ssu72-L84F and *rtr1Δ* strain have on the interactions observed with transcription proteins compared to a wild-type strain. I used Significance Analysis of Interactome (SAINT) to identify statistically significant interactions in all samples and to remove low confidence interactions using control affinity purification data [353].

METHODS

Yeast Strain Production

Yeast strains were created using the established BY4742 strain as parental and are listed in Table 2. BY4742 and the *rrp6*Δ strain were acquired from Open Biosystems. The Ssu72-L84F and *nrd1Δ*151-214 strains were acquired from Danny Reines. All cells were grown in 10 grams of yeast extract 20 grams of peptone and 100 ml of 20% glucose known as YPD or a selective synthetic drop-out media (when indicated) at 30°C. The FLAG-tagged strains were created through transformation of a cassette containing three semi-repetitive FLAG sequences, DYKDHDG-DYKDHDI-DYKDDDDK, downstream of a *URA3* gene in the pBS1539 construct (Table 3). This cassette was created with primers with homology to the 3' UTR region of *RPB3* and the pBS1539 plasmid. Stable incorporation of the FLAG tag was confirmed via Western blot using an anti-FLAG antibody (A8592 Sigma). The strains with a 3C protease cleavage site inserted into Rpb1 were made using an Agilent Quickchange Lightning Kit. The 3C cleavage sequence, LEVLFQ/GP, was inserted at amino acid 1460 of Rpb1, into the linker region of *RPB1* in an *RPB1* pENTR-D-TOPO construct and confirmed via sequencing. The construct was then transformed into a wild type strain with a 3XFLAG tagged *RPB3*. *RTR1* was knocked out of wild type and the *rrp6Δ* strain by homologous recombination with a kanamycin cassette to create the *RTR1* deletion and *rtr1Δ/rrp6Δ* double deletion strains. *rrp6Δ* strain is from the yeast knockout collection (Open Biosystems) [354]. The Rpb3-3xFLAG (referred to as Rpb3-FLAG) WT strain has been previously described [57]. The Nrd1-TAP strain is from the yeast TAP-tag collection (Open Biosystems). The Rpb3-

FLAG and Nrd1-TAP *rtr1Δ* strains were made by amplification of the *RTR1* knockout cassette from the *rtr1Δ* strain and transformation into the wild-type Rpb3-FLAG and Nrd1-TAP strains respectively. All deletion strains were confirmed by PCR-based genotyping. To perform a single biological replicate for genomics or proteomics experiments, growths of all strains of interest were pre-cultured from a single colony obtained from a sequence-verified glycerol stock of the strain that had been plated on the appropriate selective medium and grown for 2 days. Liquid cultures of all genotypes for an individual biological experiment were grown up on the same day. Cells for subsequent biological replicates were grown on different days.

Affinity Purification of Protein Complexes

Cells were grown to OD600 ≅ 1.5 in YPD broth overnight and collected by centrifugation for 10 minutes at 4000 x g, washed in H2O, and resuspended in 10 mL TAP lysis buffer (40mM Hepes-KOH, pH 7.5; 10% glycerol; 350mM NaCl; 0.1% Tween-20; fresh yeast protease inhibitors (Sigma; diluted to 1X)) per 1 gram of cell pellet. The cells were slowly transferred into liquid nitrogen using a syringe. The frozen cells were pulverized with a mortar and pestle and lysed further in a Waring Blender with dry ice. The frozen lysate was transferred to a new container and allowed to thaw at room temperature. The resulting extract was treated with 100units DNase I and 10µL of 30mg/mL heparin for 10 minutes at room temperature and clarified by centrifugation as previously described [355].

Name	Genotype	Source
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	Open Biosystems
yAM522	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 RPB3-FLAG::URA3*	This Study
yAM334	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-FLAG::URA3*	This Study
yAM423	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-FLAG::URA3 RPB1::3C*	This Study
2D1F	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ssu72*g.252A>T	D. Reines-Emory U.
yAM421	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ssu72*g.252A>T *RPB3-FLAG::URA3*	This Study
yAM486	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ssu72*g.252A>T *RPB3-FLAG::URA3 RPB1::3C*	This Study
YSB2078	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nrd1Δ151-214*	D. Reines-Emory U.
BY4741 *rrp6Δ*	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rrp6Δ*	Open Biosystems

Table 2: Yeast strains. Table of yeast strains, genotype and origin

Primer	Sequence
SSU72_L84F_FORWARD	5’- CGATCCAGCATTTGCAAGAAACC GTTCG-3’
SSU72_L84F_REV	5’-GTTACAAGTCGAACGGTTTCTTGCAAATG-3’
SSU72_L84A_FORWARD	5’-ACGATCCAGCATTTGCAATGCACCGTTC-3’
SSU72_L84A_REV	5’-CCGTTACAAGTCGAACGGTGCATTGCAA-3’
SSU72 C15S REVERSE	5’-CGCATTTCAAACTTGAAGTTTTGCACAGTTTC TGCATCAAACAACAAT-3’
SSU72 C15S FORWARD	5’- ATTGTTGTTTGATGCAGAAACTGTGCAAAAC TTCAAGTTTGAATTGCG-3’
RPB3 3X FLAG Pbs1539	5’-ATGACGGGCGCTGAGCAAGATCCTTACTCCA ATGCATCTCAAATGGGTAATACTGGATCAGGAG GGTATGATAATGCTTGGGACTACAAAGACCATG ACGGTGATTATAAAGATCATGACATCGACTACA AGGATGACGATGACAAGTGAAGCTTGATATCG AATTCCTGCAGCCCG-3’
RPB3 Pbs1539 reverse	5’-GCATATAAAGCTTTTTTTCTCTTATTATTTTCG GTTCGTTCACTTGTTTTTTTTCCTCTATTACGCCC ACTTGAGAATACGACTCACTATAGGGCGAATTG GGTACC-3’
GST_544 TO 563	5’-CGTATTGAAGCTATCCCACA-3’

Primer	Sequence
NRD1 FORWARD	5’-ATATCTGTGTGTAGAATCTTTCTCAGATATAAATTGATTGATTGAATTCGAGCTCGTTTAAAC-3’
YKL151C AS PROBE R	5’-TAAGGGACAGTCTGGTGGGAGAGTTTGC-3’
YKL151C S PROBE F	5’-CACCTCCGAGATGGGCAGCTTGATC-3’
YKL151C S PROBE R	5’-TAGCGsTAGTCTACCCAGTTTTTCAGG-3’
YKL151C AS PROBE F	5’-CACCGCTACTTGTGGAGTTATTAATTGCACTG-3’
YKL151C AS PROBE R	5’-GGTGGCTTTTTCTCAGTTTCGTGGGC-3’
YKL151C AS PROBE F	5’-CACCGCGACCTACAGCCGATGTCTTGACAGG-3’
YKL151C AS PROBE R	5’-GTTGGCCGGTACTACCGAATCACCAGG-3’
RPB1-1460 3C SITE F	5’-TTTGATGTGATGATCGATGAGGAGTCACTGGTAAAATACATGCCAGAACAAAAAATATTGGAGGTTTTGTTTCAGGGACCTACTGAGATTGAAGACGGACAAGATGGTGGCGTCACACCA-3’
RPB1-1460 3C SITE R	5’-TGGTGTGACGCCACCATCTTGTCCGTCTTCAATCTCAGTAGGTCCCTGAAACAAAACCTCCAATATTTTTTGTTCTGGCATGTATTTTACCAGTGACTCCTCATCGATCATCACATCAAA-3’

Table 3: Primers. Primers used to create strains and Northern probes

Affinity purification was performed as previously described [355]. For FLAG-tagged purifications, the lysate was incubated with anti-FLAG agarose resin (Sigma) at 4°C overnight. The resin and bound proteins were removed from the lysate by gravity flow through a 30mL Bio-Rad Econoprep column and washed on the column with 60 mL TAP lysis buffer. The resin was resuspended in 300µL of 50mM Ammonium bicarbonate pH 8.0 and transferred to a microcentrifuge tube for on-bead digestion with 5µL of Trypsin Gold (0.1µg/µL) overnight with shaking at 37°C. The supernatant containing the digested proteins was removed and treated with 20µL of 90% formic acid to inactivate the trypsin.

MudPIT-LC/MS/MS and Proteomics Data Analysis

Each affinity purified sample was loaded onto a two-phase MudPIT column containing strong cation exchange resin (Phenomenex), which binds positively charged ions, and reverse phase C18 resin (Phenomenex), which retains peptides based on their hydrophobicity [356]. The samples were eluted off the column by a MudPIT protocol of 10 steps of increasing salt concentrations (50-350mM ammonium acetate) followed by an organic gradient (20-80% acetonitrile). All chromatography solutions also contained 1% formic acid. Peptides were analyzed by a ThermoFisher LTQ Velos for MS/MS analysis. Raw spectrum data from the MS analysis were submitted for protein identification by Proteome Discoverer software (Thermo) version 2.1 using SEQUEST-HT as the database search algorithm. Database searches were performed against a FASTA database from the yeast Uniprot proteome. The FASTA database also included a number of common protein contaminants such as keratins and IgGs.

Disruption - Compensation (DisCo) Network Analysis

Disruption-Compensation (DisCo) analysis using protein-protein interaction tools to analyze protein-protein interaction dynamics as a consequence of genetic perturbation, in this case deletion of the CTD phosphatase *RTR1*. Statistical analysis of interactome (SAINT) was performed as previously described on at least four biological replicate purifications from each genotype [357-359]. In brief, PSMs for each co-purified protein were annotated per purification by bait protein, genotype (WT or 'rtr1D' for *rtr1Δ*), and replicate in the list format required for SAINT analysis on crapome.org [360]. SAINTexpress was used for the probability score calculation [358]. The output file from SAINT analysis was uploaded to ProHits-viz for prey-prey correlation analysis with the following key options: Abundance column = Spec (i.e. PSM), Score column = Saint score, Abundance cutoff for prey correlation = 20, Add bait counts = yes [361].

Global Proteome Abundance Analysis

WT and *rtr1Δ* cells were lysed in 8M urea in 100mM Tris pH 8.5 for optimal protein extraction. Samples were digested using Trypsin Gold (Promega) and labeled with Tandem Mass Tag (TMT) reagents according to the manufacturer's protocols (Thermo Fisher). TMT labeled peptide samples were combined for multiplexing, then subjected to high pH reversed phase fractionation (8 fractions). The fractions were analyzed on an Orbitrap Fusion Lumos instrument using an SPS MS3 method and the data searched on Proteome Discoverer 2.3 using a yeast proteome downloaded from Uniprot in October 2017. The protein abundances were normalized using total peptide amounts per multiplexed channel.

Transformation of Wild Type Plasmid DNA into *E. coli*

Wild type *SSU72* DNA (pGST-*SSU72* wild type supplied by the Faye lab) was purified from One Shot® chemically competent *E. coli* cells using a GeneJET Plasmid Miniprep Kit (Thermo K0503) (Table 4). An Agilent Quickchange Lightning Kit was used to conduct site directed mutagenesis of *SSU72* from TTA to TTC (L84F) and GCA (L84A) and from TGT to TCT (C15S). Rosetta 2(DES) pLys(s) cells were transformed with pGST-*SSU72*-L84A pGST-*SSU72*-L84F plasmids and pGST-Ssu72-C15S plasmids. The sequence of the mutated DNA was confirmed through third-party Sanger Sequencing (Genewiz).

RNAPII CTD 3C cleavage

Yeast cultures were grown to an OD600 of 3.5. After harvesting cells, each strain was resuspended in 10 times the cell pellet volume of TAP lysis buffer (40 mM HEPES-KOH, pH 7.5; 10% glycerol; 200 mM NaCl; 0.1% Tween-20; protease inhibitors (Sigma P8215); 1 mM sodium orthovanadate) and lysed via bead-beating. After lysis, each sample was treated with 100 units of DNase I and 0.3 μg of heparin to solubilize chromatin. Each extract was clarified by centrifugation for 1 hour at 13,000 rpm. The clarified lysate from each strain was incubated with 500 μL Anti Flag M2 (Sigma A2220) affinity resin for at least 3 hours. After incubation, each sample was applied to a Bio Rad Econo-Pac (7321010) column and allowed to drain via gravity.

Stock ID	Backbone	Plasmid	Source
pAM305	pGEX6P-1	pGST-*SSU72*-L84F	This Study
pAM306	pGEX6P-1	pGST-*SSU72*-L84A	This Study
pAM311	pGEX6P-1	pGST-*SSU72*-WT	M. Hampsey-Robert Wood Johnson
pAM312	pGEX6P-1	pGST-SSU72-C15S	M. Hampsey-Robert Wood Johnson
pAM76	pBS1539	pBS1539 with URA3 at C-terminal	Hochstrasser-Yale U.
pAM11-1	pENTR-D-TOPO	pENTR-D-TOPO RPB1	This Study

Table 4: Plasmids. Table of plasmids created and used for experiments.

The resin on the column was then washed with 100 mL of lysis buffer, followed by a wash with cleavage buffer (10 mM Tris, pH 8; 150 mM sodium chloride; 0.1% IGEPAL; 1 mM EDTA; protease inhibitors). Affinity resin from each sample was resuspended in 500 µL cleavage buffer, with 5 µL of purified 3C protease added. The RNAPII-associated proteins, affinity resin, and 3C protease mixture was then incubated overnight on a rotating wheel at 4°C. To collect the RNAPII CTD (and associated proteins), each sample was again applied to a Bio Rad Econo-Pac column, and washed with 500 µL of cleavage buffer that was collected with the CTD/associated proteins. A BCA protein assay was performed on both CTD/associated protein samples and to assist in equal protein loading onto an SDS-PAGE gel.

Whole Cell Lysate and Ser5, Ser7 Phosphorylation Western blot

Yeast strains were grown to an OD_{600} of 3.5. After harvesting cells, each strain was lysed in lysis buffer by bead-beating. Each extract was clarified by centrifugation for 1 hour at 13,000 rpm. After clarification of cell lysate, the supernatant was collected and 30 µL of each sample was boiled with Laemmli buffer. Samples were then loaded on an SDS-PAGE gel and run at 200V. Proteins were transferred to nitrocellulose paper overnight at 30V at 4°C. The membranes were then blocked with 5% milk/TBS solution for 30 minutes and incubated at 4°C overnight with one of the following primary antibodies: anti CTD-phospho-serine 2 (3E10, EMD Millipore), CTD-phospho-serine 5 (3E8, EMD Millipore), CTD-phospho-serine 7 (4E12, Millipore), or PGK1(459250, Invitrogen). Membranes were washed in TBS and incubated with their respective secondary antibodies: goat anti-rat (6845, Abcam) or donkey anti-mouse (NA934, GE

Healthcare). Finally, the membranes were developed using a Pierce ECL 2 kit and scanned on a fluorescent image analyzer (FLA-5000, FUJIFILM).

GST-SSU72 Expression in *E. coli* and Affinity Purification

The GST-SSU72 plasmid containing strains were grown in LB/ampicillin overnight at 37°C and used to inoculate 1L of LB/ampicillin, which was grown to a final OD_{600} of at least 0.8. IPTG was added to a final concentration of 1 mM at which point the incubation temperature was shifted to 18°C. After overnight protein expression, cells were pelleted and lysed three times at a pressure of ~ 1000 psi (high setting) in a French Press high pressure cell press. Lysates were clarified by centrifugation and the supernatant was loaded onto a Bio-Rad Econoprep column packed with 1 mL of Glutathione Sepharose 4B resin (GE Healthcare). Loading occurred at a rate of 1 mL/min using a Bio-Rad chromatography system. After loading, the column was washed with a wash buffer solution (50 mM HEPES-KOH, 300 mM NaCl, 10% Glycerol, 10 mM EDTA and 1 mM β-mercaptoethanol). A gradient of lysis buffer to 100% wash buffer was used at a rate of 1 mL/min for 30 minutes (lysis buffer to wash buffer). After 30 minutes, the column was washed with 100% wash buffer for another 30 minutes. After washing, the GST-tagged protein was eluted from the column using an elution buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 10 mM EDTA, 0.1-1 mM β-mercaptoethanol, 100 mM R-glutathione). Elution was conducted using an increasing gradient of elution buffer at a rate of 1 mL/min (lysis buffer to elution buffer). Elution fractions (1 mL each) were collected each minute as the UV absorbance was recorded. Fractions were combined based on the recorded absorbance and transferred to a 7000 MWCO dialysis tube.

Dialysis occurred overnight at 2-8°C in 5L of dialysis buffer (50 mM HEPES, 150 mM NaCl, 1 mM DTT). The next day, the dialysate was transferred to a 10 kDa spin filter and the solution was concentrated to ~1 mL.

In-Vitro Ssu72 Phosphatase Activity Assay

The phosphatase assay was completed in a Costar black 96 well plate. Each well was filled with water, 50 mM succinic acid, 150 mM NaCl, dialysis buffer solution, 2 μM Ssu72 protein and a final concentration of 10 μM 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP) in DMSO. After addition of DiFMUP, the plate was analyzed in a Molecular Devices Spectramax® M5 plate reader set at 30°C with excitation at 355 nm and emission at 455 nm. Glutathione S-Transferase was used as a background control (data not shown)

RNA Isolation and Northern analysis

RNA was isolated from 200 mL cultures grown to an OD_{600} of 1.0. Cells were centrifuged, washed, and then resuspended in 10 mL of AE buffer (50 mM sodium acetate at pH 5.2, 10 mM EDTA). RNA was isolated using a hot acid phenol extraction, by addition of 800 μL 20% SDS and 10ml cold acid phenol. Cells were incubated at 65°C for 10 minutes with constant agitation then incubated at 4°C. The aqueous phase was separated via centrifugation and transferred to a gel matrix phase lock tube (Qiagen), 13 mL of Chloroform was added, and RNA was separated from DNA and other impurities via centrifugation. RNA was precipitated with an equal volume of isopropanol and 1/10 volume sodium acetate. Samples were then analyzed for integrity and quality using an

Agilent Bioanalyzer by assessment of ribosomal RNA (rRNA). Afterwards the RNA was loaded onto a 1.5 % agarose gel at 30 ug total RNA and imaged with a UV imager (Bio-Rad ChemiDoc XRS). The RNA was transferred to a Zeta-Probe blotting membrane overnight by capillary action and cross-linked by UV light. The membrane was stained with methylene blue and de-stained before hybridized with radioactive DNA probes (see Table 5). DNA oligonucleotide probes were created using 5' end-labeled gamma ATP-32P by T4 Polynucleotide Kinase (Life Technologies) for detection of snRNAs. The membrane was pre-hybridized in Roche Life Science DIG Easy Hyb buffer at 37°C. The radioactive probes were added to the hybridization buffer and incubated overnight. After hybridization, the membranes were washed in 6XSSC / 0.1%SDS once at room temperature and twice for 10 minutes at 50°C. Blots were exposed to a phosphorscreen overnight for snRNAs followed by scanning using a phosphorimager (GE Healthcare). For less abundant transcript types, *YKL151C*, *IMD2* and *NRD1* primers were created to complement areas of the gene of interest where transcription was expected to occur. These primers were used to amplify a region of the gene around 250 nucleotides from the genome using PCR. A specific sequence, CACC, was incorporated into the forward primer to preserve strand specificity. The PCR product was incorporated into a pENTR/D-TOPO (Invitrogen) vector through a topoisomerase reaction and then transformed into chemically competent TOP10 cells (Invitrogen). The sequence of the PCR product in the pENTR/D-TOPO was confirmed through sequencing. The product was then transferred to a pET-DEST42 (Invitrogen) vector using LR recombination and transformed into TOP10 cells. The sequence of the PCR product in the pET-DEST42 vector was confirmed through sequencing. The PCR product and vector were isolated and

the PCR product was amplified with a T7 enzyme (MAXIscript) using P^{32} labeled UTP. The radiolabeled probe was hybridized to the RNA blot at 68˚C overnight. Next, the membranes were washed with 1xSSC/0.1%SDS twice at room temperature and twice with 0.1xSSC/0.1%SDS for 15 minutes at 68°C. Blots were exposed to a phosphorscreen overnight for 7 days followed by scanning using a phosphorimager (GE Healthcare).

RNA Sequencing and Analysis

After total RNA isolation, the RNA was DNase treated using an Ambion DNase-turbo kit. Samples were then analyzed for integrity and quality using an Agilent Bioanalyzer by assessment of ribosomal RNA (rRNA). After identifying high quality RNA, the samples were given to the Medical Genomics Core at the Indiana University School of Medicine (IUSM). At the core, 5 μg of each RNA sample were treated with the ERCC ExFold RNA Spike-In Mixes from Ambion, which included known concentrations of RNA, to all of the samples in order to control for total RNA variation. Samples were then depleted of ribosomal RNA with Ribo-Zero Magnetic Gold Kit. Afterwards cDNA libraries were prepared following the TruSeq Stranded Total RNA Sample Preparation Guide from Illumina and sequenced with the Illumina HiSeq 4000. Raw sequencing results were analyzed for quality and aligned to the yeast genome. On average each sample had ~30 million 75bp paired-end reads. The analysis was strand-specific and compared RNA expression first by normalizing according to the spike-in ratios stated in the ERCC ExFold kit and then using the edgeR program to normalize and analyze differential expression [362].

Name	Sequence
snR13 Probe	5’-GCC AAA CAG CAA CTC GAG CCA AAT GCA CTC-3’
snR3 Probe	5’-GCT CGA TCT TCG TAC TGT CTA ATG CGG TGG-3’
snR11 Probe	5’-CTA TCA ACC GCG AGC ACG ACA GTG-3’
snR82 Probe	5’-GCT TGG ATC CAA GAA AAA CTC CCG GTA ACC AGG-3’
snR33 Probe	5’-GCT AGG CTT TCA ATC TCT GCT CCT CCA AAC-3’
snR71 Probe	5’-GAT GAT AAC CTT CTC AGC TCA CTC AGA TC-3’

Table 5: RNA probes. Nucleotide sequence of probes used for Northern blots

An extended set of annotations that includes 5' and 3' UTR regions of genes as well as previously published annotations of ncRNAs, such as cryptic unstable transcripts (CUTs) and Ssu72 restricted transcripts (SRTs), were utilized, which total over 12,000 genes, to analyze differential expression [12, 13]. The edgeR program calculated normalization parameters as well as performed statistical analysis of the data using an overdispersed Poisson model, and an empirical Bayes procedure with a cut off for low read counts of 100 reads per gene [362]. These results were used for the volcano plot. Read counts per nucleotide for individual genes were calculated using the bedcov command in samtools [363]. A separate set of annotations taken from Tan-Wong *et al*. was used to compare the differential expression of the Ssu72-R129A mutant conferring the appearance of SRT's to the differential expression seen in the Ssu72-L84F mutant [13].

ChIP-exo and MNase-seq

ChIP-exo was developed to obtain high resolution data from chromatin immunoprecipitation (ChIP) followed by sequencing. This is achieved by the addition of lambda exonuclease (a dsDNA 5' to 3' exonuclease) and RecJf (a ssDNA 5' to 3' exonuclease) following annealing of the first primer set to digest the 5' end of unbound DNA and decrease the length of the DNA bound to RNAPII, thus increasing the resolution of the sequences at which RNAPII is bound on the genome (Figure 6) [20, 352].

ChIP-exo was performed as previously described [20, 352]. Briefly cells were grown in a 200 mL YPD culture to an OD_{600} of 0.8 and crosslinked with formaldehyde

(1%) (Sigma F8775– 25ML) for 15 minutes. Cells were then collected and lysed, and chromatin was sonicated to attain DNA fragments of 100-500 base pairs and visualized with the Agilent Bioanalyzer. Immunoprecipitation was done with 50 µL of anti-FLAG agarose resin. The chromatin was then barcoded with DNA adaptors and treated with lambda exonuclease and RecJf, thereby digesting the 5' end of unbound DNA. Chromatin was then un-crosslinked, isolated using chloroform and precipitated with ethanol.

Micrococcal nuclease (MNase) digestion and sequencing was performed through adaptation of the protocol by Wal and Pugh [364]. Following optimization of the digestion conditions, 15U of MNase was added to a chromatin slurry and incubated with shaking at 37 °C for 20 minutes. The digestion was quenched by addition of 50 mM EDTA and 0.2% SDS. The digested DNA was cleaned up through phenol/chloroform extraction followed by ethanol precipitation with 20ug of glycogen (Sigma) as a carrier.

ChIP-exo and MNase library construction, EZBead preparation, and Next-Gen sequencing were completed using standard methods based on the Life Technologies SOLiD5500xl system as previously described.

ChIP-exo Sequencing and Analysis

Samples were sent to the IUSM Medical Genomics Core for sequencing. EZBead preparation and sequencing were performed on Life Technologies SOLiD5000xl system as specified previously [20, 248]. Raw sequencing results, with an average of 40 million 75bp single-read reads per sample, were analyzed for quality and aligned to the yeast genome, SacCer3, by the genomics core.

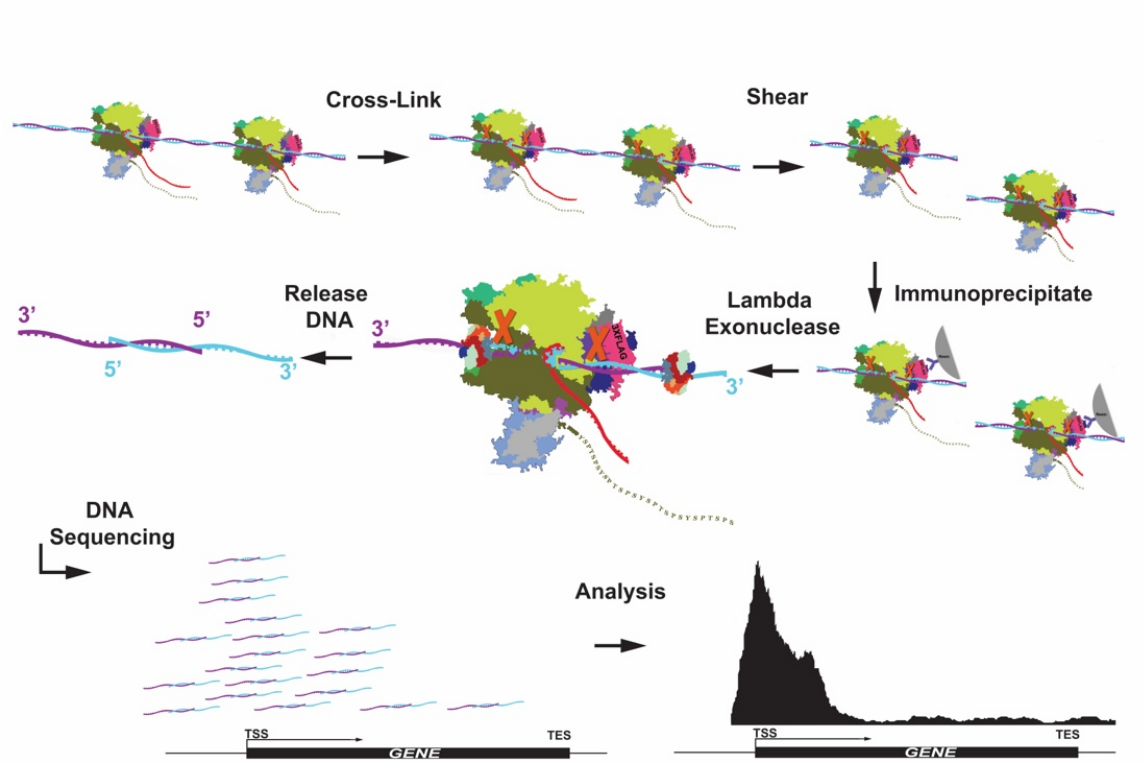


Figure 6: Chromatin immunoprecipitation and exonuclease analysis (ChIP-exo) workflow. RNAPII is immunoprecipitated, chromatin is sheared and double stranded DNA has 5' ends digested by Lambda exonuclease up to RNAPII creating a footprint up to nucleotide resolution of where RNAPII is located on the DNA.

Aligned results were then analyzed for differential localization of each gene using EdgeR similar to the RNA differential expression analysis. Average gene analysis was done using NGS.plot [365]. Different transcript types were grouped and averaged using NGS.plot for each gene for that particular group along the window specified. Before averaging read counts, all reads from each sample were normalized to each other by total read counts of each sample. Read counts per nucleotide for individual genes were calculated using the count command in the bamutils tool which is a part of the ngsutils package [366].

Cross-linking Mass Spectrometry

Cells were grown to OD600 ≅ 2.5 in YPD broth overnight. The cells were then collected by centrifugation for 5 minutes at 3000 x g, washed with H_2O, and pelleted via centrifugation at 3000 x g for 5 minutes before being stored in 50 mL falcon tubes at -80°C. Frozen cell pellets were thawed and transferred to liquid nitrogen dropwise using a syringe. The frozen cell droplets underwent cryogenic grinding using the Cryomill (Restch). TAP lysis buffer (40 mM HEPES -KOH, pH 7.5; 10% glycerol; 350 mM NaCl; 0.1% Tween-20) and yeast protease inhibitor (Sigma; diluted to 1X) were added to cells at room temperature. Cells were centrifuged at 14,000 x g for 1 hour. The lysate was then transferred to a new container and incubated with TAP lysis buffer washed anti-FLAG agarose resin (Sigma) at 4°C overnight. The resin and associated proteins were then removed from the lysate by gravity flow through a 30 mL Bio-Rad Econoprep column. The column was then washed with 120 mL of TAP lysis buffer followed by two additional washes with 500 µL of 20 µL HEPES buffer. During the second wash, the HEPES buffer was retained in the column so that 2 mg of disuccinimidyl sulfoxide

(DSSO) dissolved in 102 µL could be added to the protein solution. The solution was incubated on the column at room temperature for one hour. 10 µL of tris was added and the resin and associated proteins were extracted with 200 µL of 8 M urea prior to transfer to a 1.5 mL microcentrifuge tube. 10 µL of TCEP was then added to the protein sample which was then left to incubate at room temperature for 30 minutes. This was followed by an addition of 3 µL CAM and an incubation period of 30 minutes in the dark. 2 µL of LysC (.2µg/µl) was added to the sample which and incubated at 37°C overnight with shaking at 900 rpm. 600 µL of 100 mM Tris buffer was added to the sample, followed by 4 µL of 1 M $CaCl_2$. 10 µL of Trypsin Gold (.1µg/µl)) was added to the sample and incubated overnight at 37°C with shaking at 900 rpm.

RESULTS

Rtr1 Regulates Premature Transcription Termination

The production of mRNA is very well studied and uses many complexes of proteins to make a stable transcript that will go on to be translated. However, ncRNA is more difficult to study and there is more limited knowledge of how it is regulated through RNAPII. In this study, I discovered a possible role for Rtr1 in the regulation of both ncRNA and mRNA production. I used an *RTR1* knockout yeast train to study the effects of Rtr1 on RNAPII transcription as well as the changes in interactions that it causes in the transcription machinery. My studies provide evidence that premature termination occurs in *RTR1* knockout cells through the NNS complex subunits Nrd1 and Nab3. I also used an *RTR1 RRP6* double knockout strain to compare the *rtr1Δ* cells in order to isolate specific changes in the mechanism that do not involve RNA degradation. This work suggests that one of the functions of Rtr1 in the cell is to promote elongation and that it is a negative regulator of the NNS termination pathway.

1. *RTR1* Deletion Results in Changes in PPI Network of Termination Factors

Phosphorylation of RNAPII CTD is an important regulatory mechanism through which transcription termination is directed [237, 367-369]. Deletion of *RTR1* leads to a global increase in the phosphorylation of Ser-5 of the CTD, and previous studies have shown that loss of *RTR1* results in 3'-end processing defects at the polyA-dependent gene *NRD1* [203]. In order to determine the overall role of Rtr1 and how it might regulate RNAPII and its PPIs, I performed a DisCo network analysis. For this analysis I used affinity purification mass spectrometry (AP-MS), where a tagged protein is purified along

with interacting proteins and identify the interacting partners with a mass spectrometer, with different strains, including Nrd1-TAP, Pcf11-TAP, and Ssu72-FLAG, in combination with the *RTR1* deletion to assess the effect of Rtr1 on different facets of transcription. DisCo utilizes AP-MS to analyze genetic perturbations to get more in depth and large-scale analysis of the cell and understanding of the mechanisms involved. It is possible that small changes in specific genes have an effect on PPI networks at non-primary interactions. In this study, I used SAINT probability scores from yeast cells with an *RTR1* deletion to create a protein-protein interaction network [357, 358]. This experiment resulted in a large data array consisting of 24 x 3,960 protein-level measurements from 668,969 peptide-spectrum matches. In order to gain a more realistic understanding of the effect of the mutation, I used high-confidence interactions in the termination factor complexes as well as Rpb1 and Rpb2 of the RNAPII complex.

Purifications of Nrd1-TAP, Pcf11-TAP, and Ssu72-FLAG in WT and *RTR1* deletion genotypes were performed and analyzed using prey-prey correlation analysis. Proteins with high correlation values have a similar pattern of PSMs across different sets of purifications whether or not the same bait protein is used for the purification. This is especially true with proteins that act together in the same protein complex or mechanism and often have the highest correlation values (Figure 7A). My data show that there is an association between all of the CF1A members and CPF members Fip1 and Pap1. The CPF members are components of the poly(A) polymerase module, and this association reveals that there is a cross-association between CF1A and the CPF complex specifically with the poly(A) polymerase module. (Figure 7A) [370]. In addition to the DisCo analysis, I also used SAINT to look at individual purifications of the *RTR1* deletion

versus WT with different tagged proteins Ssu72, Nrd1 and Pcf11. Both the DisCo results and the SAINT results reveal that there is an association between CPF and CF1A and that Rtr1 may regulate some of these interactions (Figure 7C). The data supports a previous hypothesis that linked the CID of Pcf11 to the formation of a CPF/CF1A/RNAPII complex necessary to form a stable RNAPII complex at the 3' end of the gene during mRNA polyadenylation [371]. This is evident in the DisCo data when comparing the WT to the *RTR1* deletion strain. Loss of Rtr1 results in a significantly decreased, if not eliminated, association between the CF1A complex proteins and the CPF members (Figure 7B). These data also show a reduction in the correlation between CF1A and RNAPII subunits Rpb1/2. As SAINT analysis contains a probability algorithm, additional information can be gained about the protein interactions among the termination machinery or transcription proteins. Ssu72 is a phosphatase and an essential member of the CPF complex. All CPF subunits had SAINT probabilities of ≥ 0.95 (Figure 7C). This is consistent with previous work from the Mosley lab showing that SAINT probability scores of ~0.95-1 are indicative of proteins within the same complex. The SAINT scores for proteins that are not in complex with the bait protein can vary greatly depending on the degree to which they interact with the bait. This variation can be decreased through replicate experiments and by increasing specificity. Data from the Mosley lab has shown that proteins known to interact with RNAPII in a more dynamic fashion can have a range of SAINT probabilities, for example 0.23 for Spt16 or 1 for Spt5, Pta1, Tfg1, and Tfg2, in replicate Rpb3-TAP purifications [372]. Without prior knowledge of a known PPI, a low interaction score would reveal that this interaction does not have a high probability of being a true interaction. However, it is possible that the interaction is not always present

in the cell or is not always detectable for a variety of reasons including stability, low affinity or heterogeneity of the interactions in a complex or in the cell. In the Ssu72-FLAG purifications, the two largest subunits of RNAPII, Rpb1 and Rpb2, were identified as significant interacting proteins with Ssu72 in WT and *rtr1Δ* cells while *rtr1Δ* cells had a significant interaction between Ssu72 and Rpb3 (Figure 7C). Rpb1 and Rpb2 are the largest proteins in the RNAPII complex and contain the most detectable peptides amongst the twelve RNAPII subunits and therefore the highest probability of detection when using AP-MS [355]. These data show that the loss of Rtr1 leads to a change in the interactions in the other serine-5 phosphatase Ssu72 and its complex with RNAPII specifically by increasing CPF interactions with Rpb1 and Rpb2 (Figure 7).

SAINT analysis of the CF1A tagged protein Pcf11-TAP and Pcf11-FLAG purifications from WT and *RTR1* deletion cells ($n \geq 4$) detected the four protein members of CF1A with probabilities ≥ 0.95, which supports their designation as a protein complex (Figure 7C). Previous studies have shown that the purified Pcf11 CID domain has lower affinity for Ser5-P and Ser2,5-P than Ser2-P modified CTD [371]. The Mosley lab has shown, using ChIP assays with antibodies specific for S2P and S5P, that Ser2/5 is less phosphorylated on genes globally as a result of *RTR1* deletion in the cell and histone H3K36me3 levels are increased which would be expected with less Ser2/5 phosphorylation [373]. In WT cells, Rpb1 and Rpb2 have a high SAINT probability for interacting with Pcf11. In the *RTR1* deletion cells the interaction between the RNAPII subunits and Pcf11 disappears statistically in both the SAINT analysis and the prey-prey correlation analysis (Figure 7).

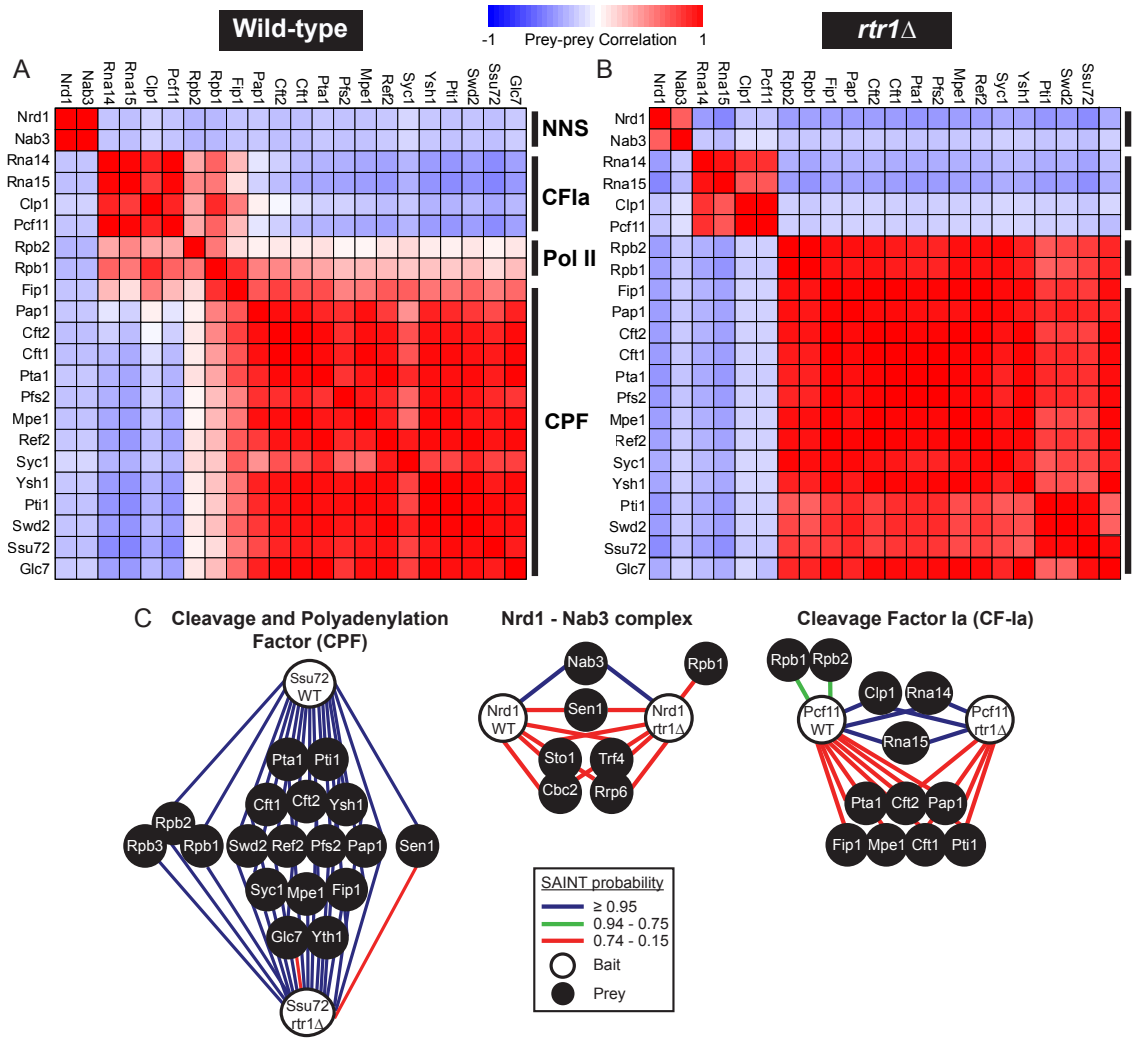


Figure 7: DisCo network analysis of RNAPII transcription in WT and *rtr1Δ* cells in yeast. (A) Prey-prey correlation analysis of the termination complex AP-MS data from WT cells of ≥ 4 biological replicate purifications for bait proteins (Nrd1, Pcf11, and Ssu72, total n=12). (B) Prey-prey correlation analysis of yeast termination complex AP-MS data from *rtr1Δ* cells of ≥ 4 biological replicate purifications for each bait protein (Nrd1, Pcf11, and Ssu72, total n=12). (C) Ssu72-FLAG, Pcf11-TAP and Pcf11-FLAG, and Nrd1-TAP protein-protein interaction networks from WT and *rtr1Δ* cells depicting SAINT analysis of ≥ 4 biological replicate purifications for each genotype. The nodes represent a protein of interest whereas the edges represent the SAINT interaction probability as indicated in the legend.

Although the data show that there are no interactions between Pcf11 and Rpb1, it is still a plausible interaction, but it is below the detection limit in this particular AP-MS study.

Many of the CPF subunits were found in the AP-MS study using the Pcf11-FLAG at low levels and were more apparent in the SAINT analysis relative to the correlation analysis. Of interest, in the *RTR1* deletion cells, the PPI SAINT scores for CPF proteins in the CF1A Pcf11-FLAG purification was not increased nor decreased relative to the WT purifications but did have less CPF subunits (Figure 7C). Thus, it is possible that CF1A still interacts with the CPF subunits, even with a decrease of Pcf11 and RNAPII interactions, due to the loss of Rtr1. It is likely that the interactions between CPF and CF1A are maintained due to their role in regulating nascent RNA and are stabilized through their interactions with a phosphorylated RNAPII CTD in a WT cell.

Nrd1 has been reported to function in complex within both the NNS complex (Nab3, Sen1) and the with capping complex proteins Cbc2, and Sto1 [374]. My AP-MS SAINT analysis of an Nrd1-TAP tagged protein does not support this theory. Only the Nab3 protein co-purified with Nrd1 well enough to have high SAINT scores of ≥ 0.95, a result consistent with the pre-prey correlation analysis (Figure 7). Although Sen1, Sto1, and Cbc2 were identified as Nrd1 interacting proteins, their SAINT probability values resemble those of dynamic or low affinity interacting partners of the bait Nrd1 and not of proteins that would be in complex with Nrd1 (Figure 7C). In addition, subunits of the TRAMP complex and the nuclear exosome were detected as dynamic interacting partners of Nrd1 determined by SAINT analysis, consistent with work from the Buratowski lab (Figure 7C) [374]. Previous work has shown that the Nrd1 CID has the highest affinity for Ser5-P and Ser2,5-P CTD repeats and the abundance of these phosphorylations are

increased in *RTR1* deletion mutants [203, 368, 373]. Nrd1 also had an interaction with Rpb1 and was identified with a SAINT score of 0.49 in *rtr1Δ* cells. Taken altogether these data show that loss of Rtr1 increases the interaction between Nrd1/Nab3 heterodimer and RNAPII *in vivo*. This is most likely due to the increase in phosphorylation of serine 5 of the CTD. In *RTR1* deletion cells the probability score of the interaction between Rpb1 and Nrd1 was lower that what was calculated for Pcf11 from WT cells. Thus, it is possible that the interaction of Rpb1 and Nrd1 occurs, but at a lower frequency than the interaction between Rpb1 and Pcf11.

2. Rtr1 Regulation of ncRNA Gene Expression

In order to analyze how the transcriptome changed after loss of the S5P phosphatase Rtr1, I performed strand-specific RNA-seq of total RNA in a set of four biological replicates. Since it was expected that *RTR1* deletion would have a global effect on transcription, this experiment required an external standard, such as a spike-in of a known quantity of RNA, to be able to measure the changes in the transcriptome [375]. RNA reads were aligned and then analyzed for differential expression using the edgeR package (Figure 8) [362]). In order to identify differential expression of certain classes of genes such as sn/snoRNAs, ORF-Ts, CUTs, SUTs, SRTs and NUTs, I used annotations that were previously described from work from several labs [13, 160, 376]. I also used annotations from previous work from the Mosley lab where RNAs that are antisense to ORF-Ts are labeled as ASTs [377]. In this study 1481 out of 11,151 transcripts were reduced in *rtr1Δ* cells, many of which were ASTs as well as other ncRNA (Figure 8A/B).

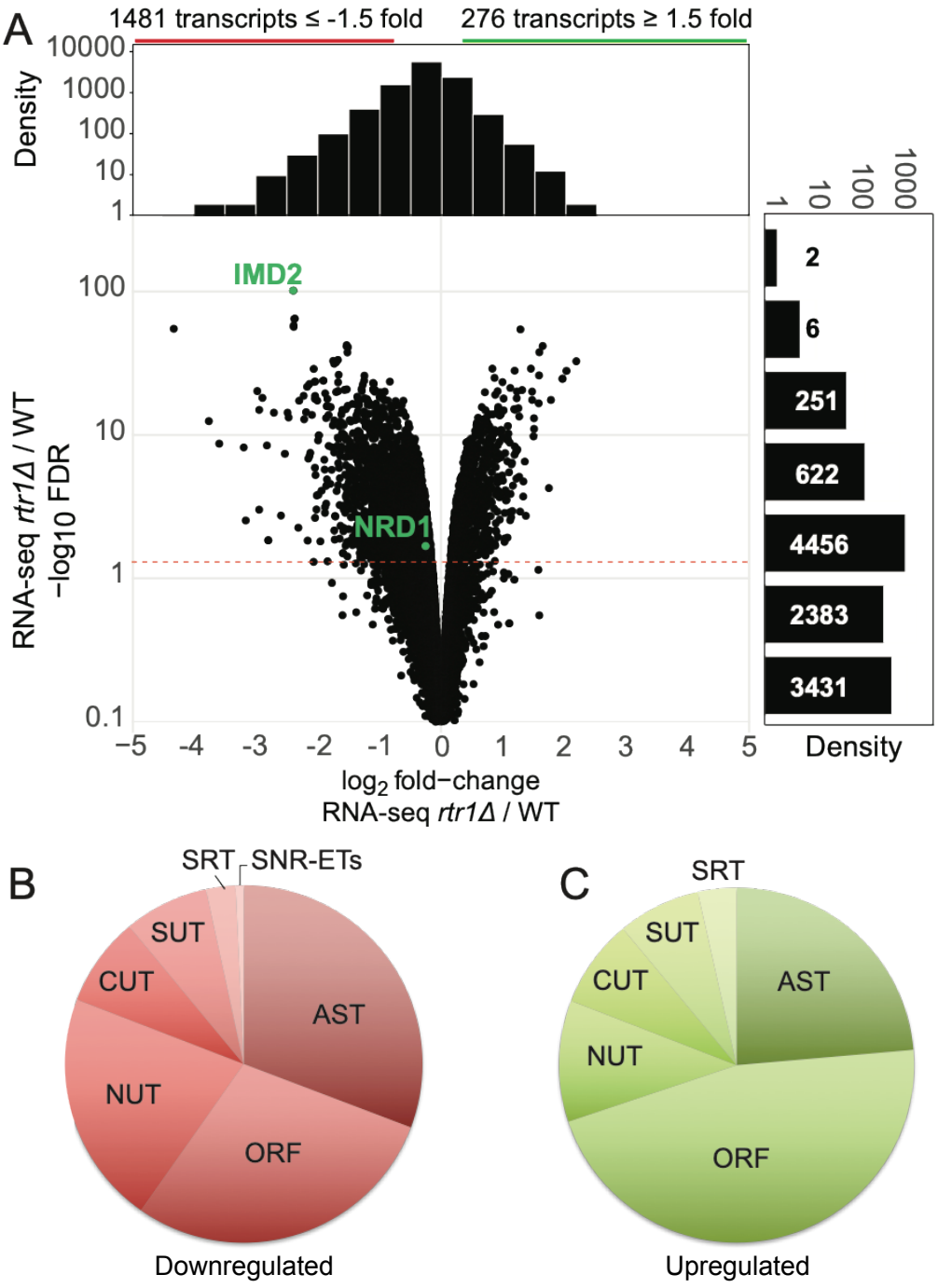

A
1481 transcripts ≤ -1.5 fold
276 transcripts ≥ 1.5 fold
Density
10000
1000
100
10
1
IMD2
NRD1
RNA-seq rtr1Δ / WT
−log10 FDR
100
10
1
0.1
−5 -4 -3 -2 -1 0 1 2 3 4 5
log2 fold−change
RNA-seq rtr1Δ / WT
1 10 100 1000
2
6
251
622
4456
2383
3431
Density
B
SRT
SNR-ETs
SUT
CUT
AST
NUT
ORF
Downregulated
C
SRT
SUT
CUT
AST
NUT
ORF
Upregulated

Figure 8: Loss of Rtr1 activity leads to global differential expression changes in ncRNA in yeast. **(**A) Volcano plot of normalized RNA-Seq data for protein coding genes and non-protein coding genes in *RTR1* deletion vs. WT cells (n=4, edgeR analysis). Data includes all annotated ORF-Ts cryptic unstable transcripts (CUTs), Nrd1 unterminated transcripts (NUTs), SNR extended transcripts, Ssu72 restricted transcripts (SRTs), and antisense transcripts (ASTs). False discovery rate (FDR) cutoff of ≤ 0.05 is apparent with a dashed red line. Plots also show the quantity of genes that correspond to the number of genes that are at each axis for comparison. (B&C) Pie charts of transcripts differentially expressed in *rtr1Δ* cells relative to WT cells across different RNA transcript classes. Downregulated transcripts are shown in the red diagram while upregulated transcripts are shown in the green diagram.

Fewer transcripts were upregulated, 276 were increased more than 1.5-fold, most of which were ORF genes (Figure 8C). Of note, the *IMD2* gene had the biggest reduction of expression while its upstream CUT was also significantly reduced. This gene is well described and is known to be regulated by the NNS termination pathway through an NNS terminator signal, but is also regulated by GTP levels in the cell (Figure 8A, labeled in green) [242, 351, 378-380]). My data interestingly show that somehow Rtr1 is helping to regulate both of the IMD2 and IMD2 CUT genes possibly independently through the NNS pathway.

Overall, the data show that loss of Rtr1 phosphatase activity results in the downregulation of different classes of ncRNAs and also has an effect on specific mRNA genes. It is possible that the increased levels of Serine 5 phosphorylation on the CTD of RNAPII in the *RTR1* deletion cells may cause increased activity and/or efficiency of the NNS termination mechanism. Increasing the efficiency of the NNS mechanism would lead to increasing the rate of premature transcription termination of ncRNA and possibly some mRNA.

3. RNAPII and Nrd1 Occupancy are Reduced at SNR Genes in *rtr1Δ* Cells

Analysis of total RNA in the *RTR1* deletion cells revealed that many ncRNA classes had a reduced expression including SNR genes (Figure 8). SNR extended transcripts (ETs) in particular, which are genes that have been shown to transcribe past their termination site when termination is defective, had a lot of differential expression [381]. SNR transcripts generally need to be processed at their 3' end by NNS termination in conjunction with the TRAMP complex and finally the RNA exosome [374, 381-386]. I performed an average gene analysis using ChIP-exo datasets for both Rpb3 and Nrd1 for the SNR genes aligned to the TSS with 500 bp upstream and 1000 bp downstream of the TSS [365]. The average RNAPII occupancy at SNR genes was decreased in *RTR1* deletion cells (Figure 9A). The Nrd1 signal at the SNR genes was also decreased. In addition, Nrd1 levels decreased to a greater degree than Rpb1 proportionally, which might suggest that lower levels of Nrd1 are necessary for termination in cells without Rtr1 or that efficiency of Nrd1 has increased as a result of the deletion of *RTR1*. Also, the reduction of RNAPII levels at SNR genes could mean that loss of Rtr1 may be affecting total RNAPII that are undergoing transcription by increasing the speed of termination (Figure 9). This hypothesis is bolstered further by the data that show that the SNR-ETs are decreased in the *RTR1* deletion cells (Figure 9B). This could suggest that termination is occurring earlier than normal at these SNR genes due to the lack of Rtr1. The argument could be made that overall transcription for the SNR genes is reduced. However, my data show that this is not the case since several mature SNR transcripts including snR62, snR32 snR189 and snR46 are not statistically any different in the *rtr1Δ* cells vs the WT cells which would mean that the effects are only occurring at the 3' end of the genes.

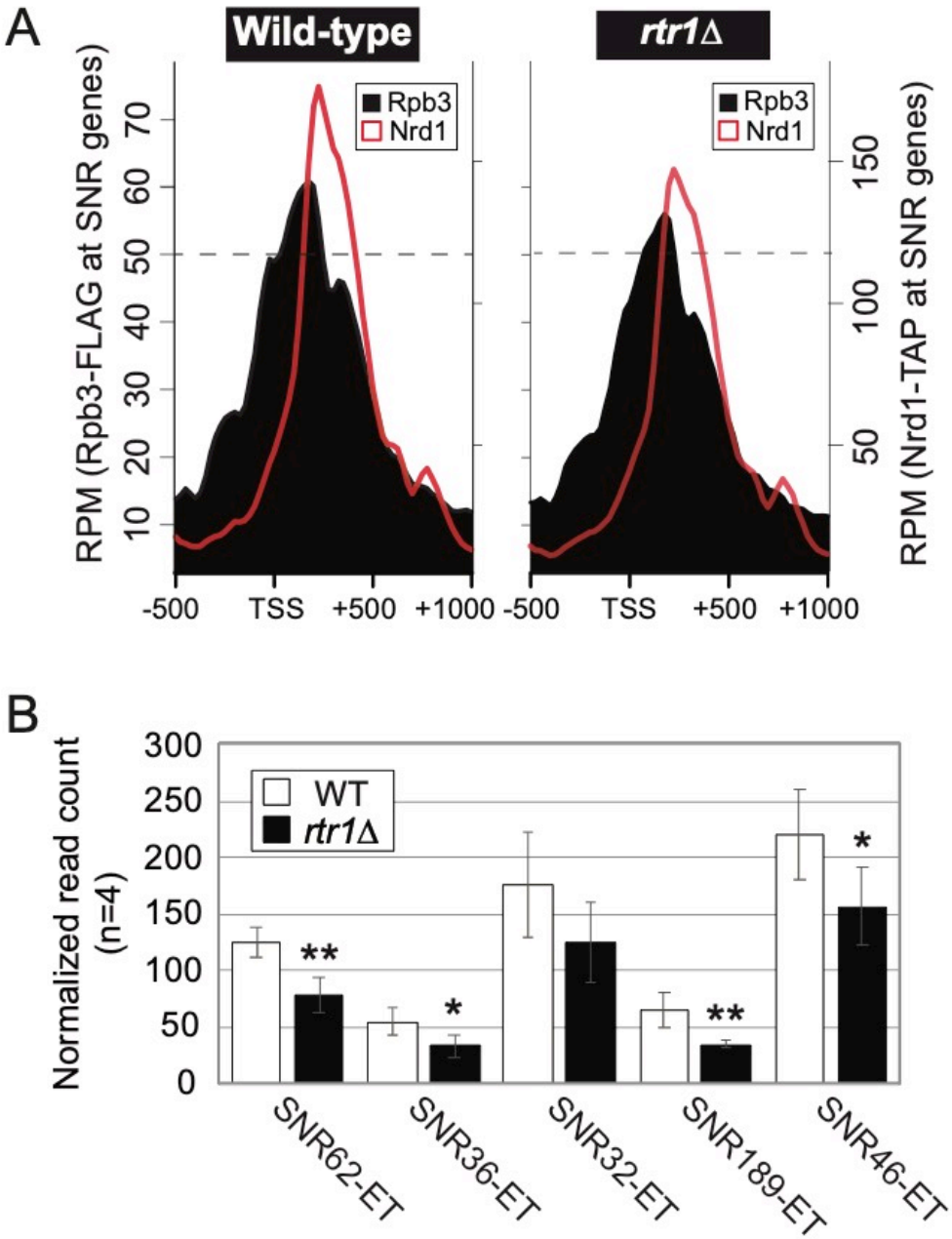


Figure 9: Loss of Rtr1 phosphatase activity leads to decreased RNAPII signal at SNR genes and leads to shortened SNR transcripts. **(**A) Average gene analysis of RNAPII and Nrd1 signal at SNR genes in WT or *RTR1* deletion cells. The vertical axis values are average read counts per million (RPM) values calculated using ngs.plot -500 and +1000 nucleotides relative to the SNR gene annotated TSSs in either WT or *RTR1* deletion cells. (B) RNA-Seq read counts for manually annotated SNR-ET regions from a subset of SNR genes.

4. Global Levels of RNAPII and Nrd1 Occupancy are Transformed in *rtr1Δ* Cells

Nrd1 localization is highest at genes transcribed by RNAPII that contain the RNA binding motif for Nrd1 and Nab3, such as the *URA8* gene. Others have previously shown that Nrd1 accumulates on the gene downstream of the peak of serine 5 phosphorylation of the RNAPII CTD of protein coding genes, a result which I have also replicated in this study [149, 151]. In contrast, protein-coding gene *PMA1* RNAPII occupancy is similar across the entire gene, but Nrd1 localization is higher at the 5' end of the gene ~270-321 bp past *PMA1*s transcription start site (data not shown). In order to see the effect that Rtr1 has on RNAPII and Nrd1 localization, I compared the average occupancy of Nrd1-TAP and Rpb3-FLAG at 2426 protein coding genes looking at a window that starts at 200 bp upstream of the genes TSS and 1000 bp downstream of the TSS and overlaid the data to average histone occupancy data in WT cells from an MNase-Seq experiment(Figure 10). In addition, I plotted the same data at a different window of 200bp upstream of the TES and 200bp downstream of the TES. Occupancy of Rpb3 was higher at the 5' end of the genes in the *rtr1Δ* cells compared to the WT. In contrast Rpb3 localization decreases in the *rtr1Δ cells* more than in the WT cells at the 3' end of the protein coding genes as well as at the TES (Figure 10A). Nrd1 occupancy also displayed decreased occupancy at the 3' end of the gene in the *rtr1Δ* strains relative to WT (Figure 10B). In addition, the ratio of RNAPII occupancy at the TSS relative to the TES at genes of varying occupancy levels further displays that deletion of *RTR1* leads to a decrease in RNAPII occupancy at the TES (Figure 10C). When the Rpb3-FLAG and Nrd1-TAP ChIP-exo average global

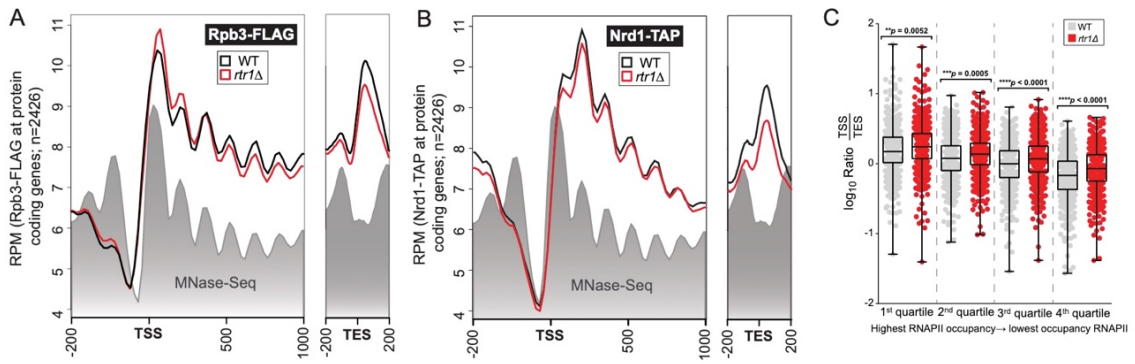


Figure 10: *RTR1* deletion leads to decreases RNAPII occupancy at the transcription termination site of protein-coding genes. Average protein coding gene analysis of RNAPII (Rpb3) (A) or Nrd1 (B) levels at 2,426 protein-coding genes -200 bps upstream and +1000 downstream of the TSS and -200/+200 upstream/downstream of the TES. The data on the vertical axis is average RPM values calculated using ngs.plot for each genotype designated by the colors in the legend. (C) Log ratio of RNAPII occupancy at TSS relative to TES at four quartiles separated by occupancy.

data at protein coding genes are overlaid on each other it is clear that the RNAPII occupancy begins to decrease just after Nrd1's highest localization peak (Figure 11).

IMD2 is a well-studied gene that has an upstream CUT transcript. The *IMD2* CUT gene is terminated by the NNS pathway but is usually transcribed and decreases the levels of *IMD2* mRNA. *IMD2* mRNA levels are low while levels of the guanine nucleotide are high, but when the guanine levels are low *IMD2* CUT is terminated early due to the fact that an alternative TSS is used and *IMD2* mRNA levels increase in the cell to produce more guanine (Figure 12A). As these cells were grown in YPD, there was sufficient guanine so that *IMD2* CUT should be produced at an earlier TSS and terminated at the intergenic terminator by the NNS termination pathway (Figure 12A). When guanine is low the TSS that RNAPII utilizes is within the intergenic terminator. (Jenks Reines 2008). Using ChIP-exo, one can analyze RNAPII and Nrd1 localization, the results showing that both RNAPII and Nrd1 mapped to the upstream region of *IMD2* at the CUT region with low levels of RNAPII localized at the protein coding region in WT cells (Figure 12B/C). In *rtr1Δ* cells, both Nrd1 and RNAPII levels were reduced just as the SNR genes were reduced (Figure 12B/C). Additionally, I performed RNA-seq analysis of *rrp6Δ* and *rtr1Δ rrp6Δ* cells in order to analyze the role that the RNA exosome might have on ncRNA in cells lacking Rtr1 (Figure 13A/B).

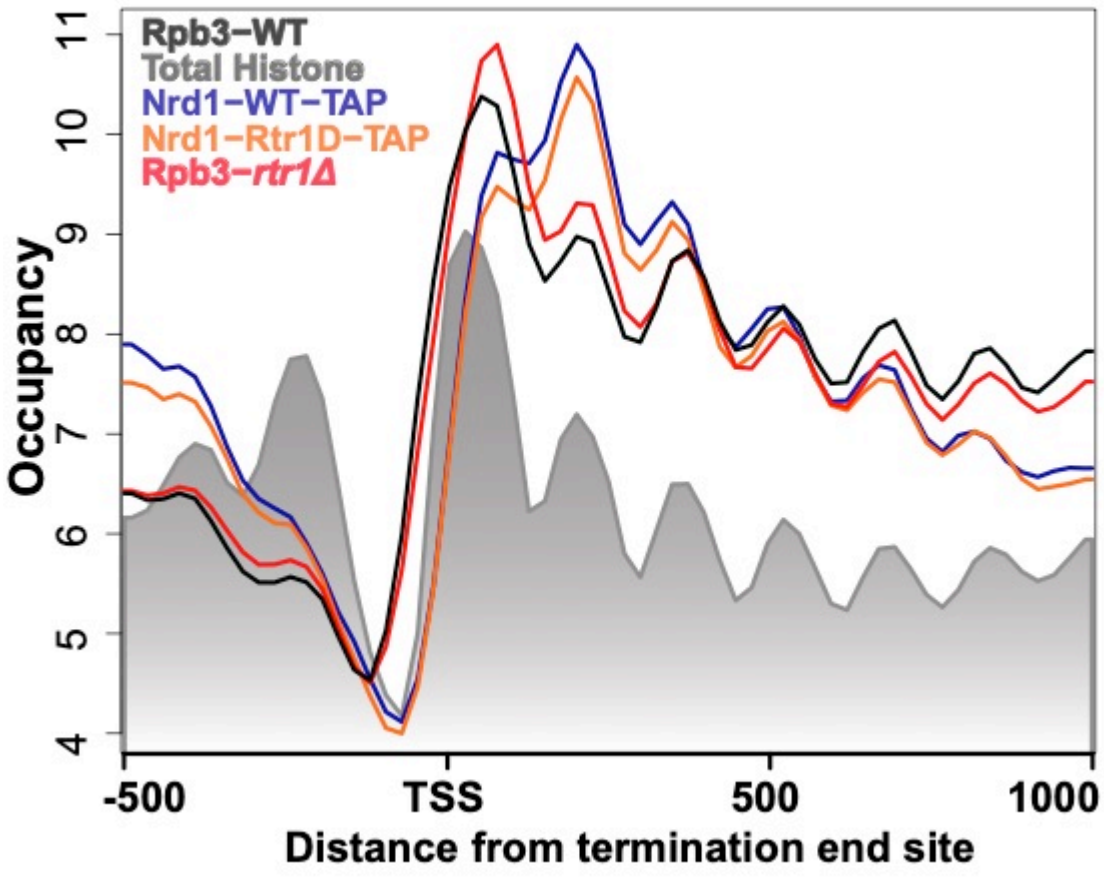


Figure 11: Gene average profiles of ChIP-exo occupancy for RNAPII (Rpb3) and Nrd1 in WT and RTR1 knockout cells. The legend displays the color for each sample. MNase-Seq based histone occupancy is also shown as gray shaded profiles [365].

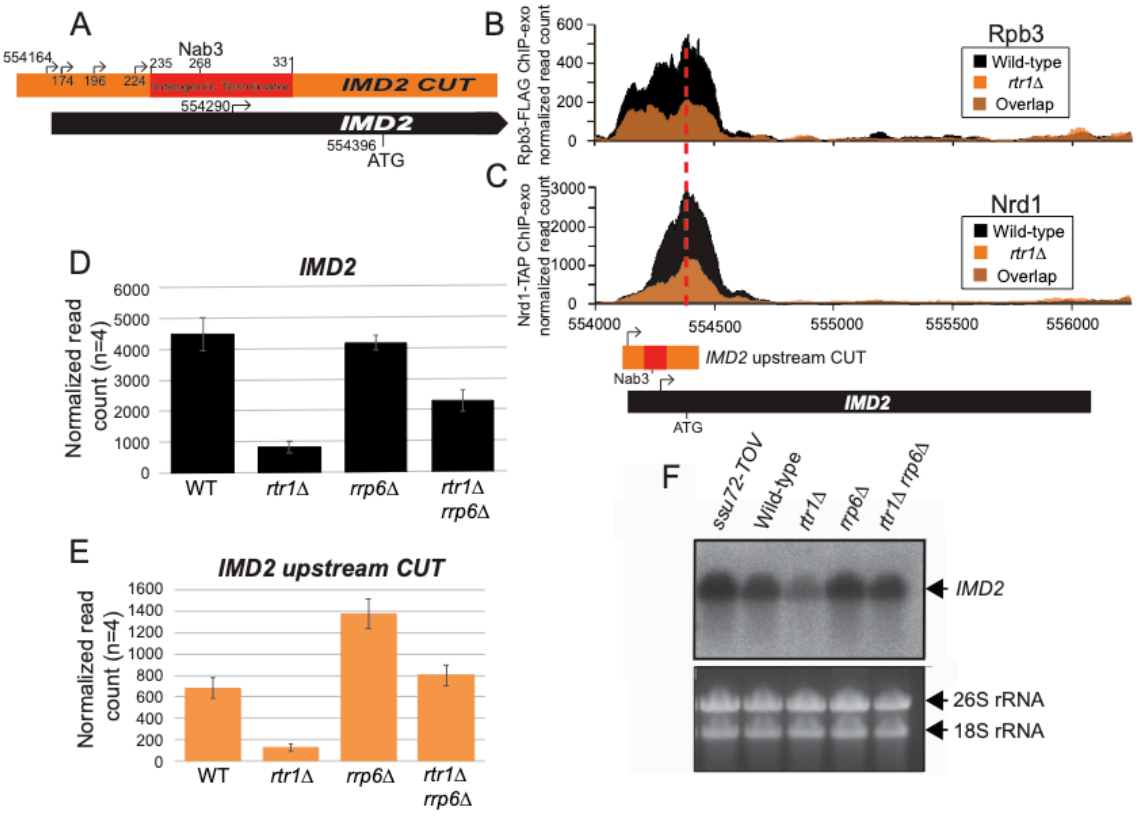
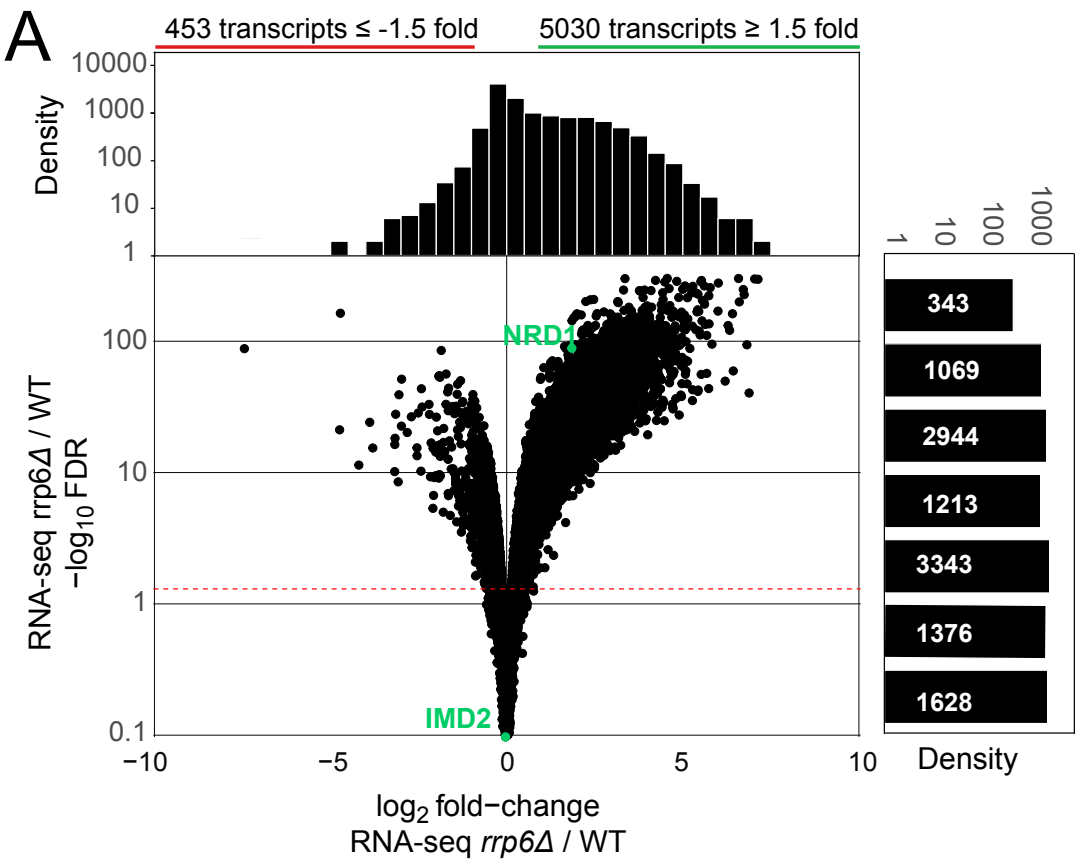


Figure 12: Rtr1 is required for basal *IMD2* expression. (A) Schematic representation of NNS-dependent attenuation of the *IMD2* gene. Arrows designate alternative start sites. Occupancy of RNAPII (B) and Nrd1 (C) at the *IMD2* gene. Data derived from WT cells is in black, and those from the *rtr1Δ* strain are in orange. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs, each to scale. Red dashed line represents TSS of IMD2. D) Expression of the *IMD2* mRNA RNA-Seq (n=4). E) Expression of *IMD2* upstream CUT by RNA-Seq (n=4). F) Northern blot analysis of *IMD2* expression in various genotype cells as indicated


A
453 transcripts ≤ -1.5 fold
5030 transcripts ≥ 1.5 fold
Density
10000
1000
100
10
1
RNA-seq rrp6Δ / WT
−log10 FDR
100
10
1
0.1
NRD1
IMD2
−10
−5
0
5
10
log2 fold−change
RNA-seq rrp6Δ / WT
1
10
100
1000
343
1069
2944
1213
3343
1376
1628
Density

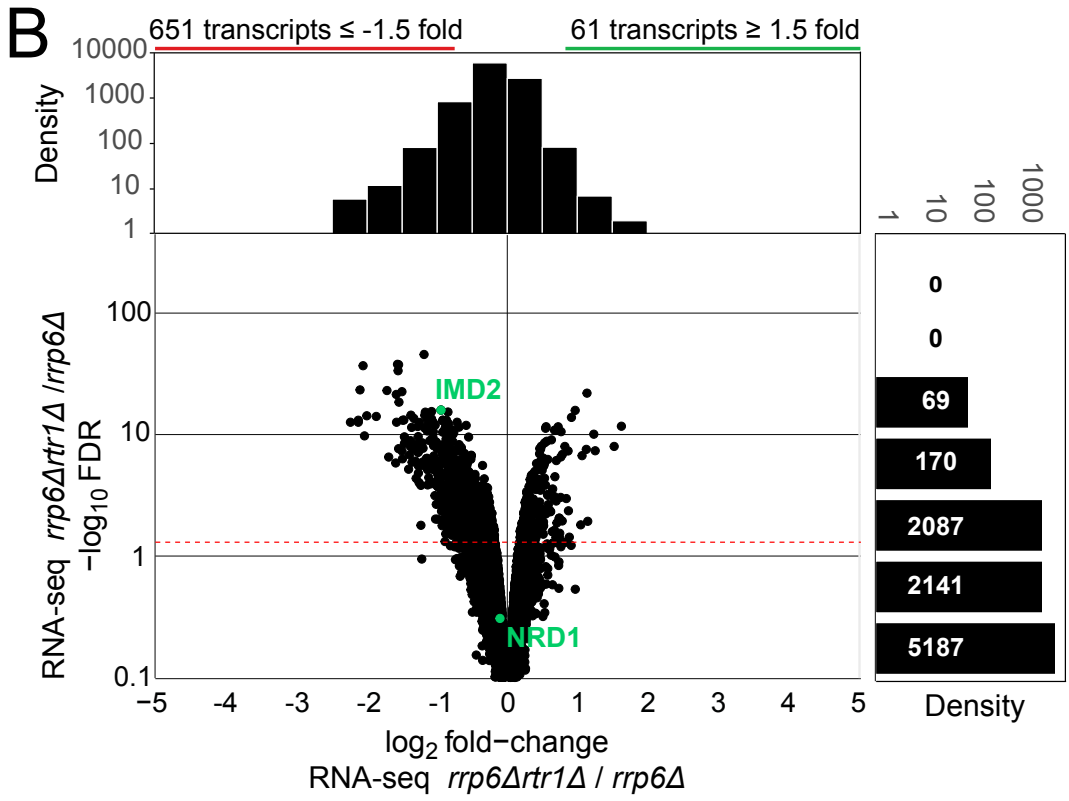

B
651 transcripts ≤ -1.5 fold
61 transcripts ≥ 1.5 fold
Density
10000
1000
100
10
1
RNA-seq rrp6Δtr1Δ /rrp6Δ
−log10 FDR
100
10
1
0.1
IMD2
NRD1
−5
-4
-3
-2
-1
0
1
2
3
4
5
log2 fold−change
RNA-seq rrp6Δtr1Δ / rrp6Δ
1
10
100
1000
0
0
69
170
2087
2141
5187
Density

Figure 13: Volcano plots in *RRP6* and *RRP6 RTR1* knockout transcriptomes relative to WT. Density plots are on the right and illustrate the number of points in each area as indicated. The number of decreased and increased transcripts based on a fold-change cutoff of 1.5-fold and an FDR of at least 0.05 are shown at the top of each panel for *rrp6Δ* (A) and *rrp6Δ rtr1Δ* (B) [387].

Previous studies have shown that by deleting *RRP6* the upstream *IMD2* CUT is increased, which my results confirmed, although the CUT was decreased in the *rtr1Δ* strain (Figure 12) (Davis Ares 2006 and Steinmetz Brow 2006). The *rtr1Δ rrp6Δ* cells maintained steady levels of the upstream *IMD2* CUT relative to WT indicating that the RNA exosome might be degrading the CUT RNA following NNS termination in cells lacking Rtr1. It is also possible that there could be change in the elongation mechanism that could have had an effect on the differential levels of the CUT RNA in the *rtr1Δ* and *rtr1Δ/rrp6Δ* cells. Interestingly, the results show that the *IMD2* coding region is altered in the *rtr1Δ* probably because it is close to the Nrd1/Nab3 binding sites of the *IMD2* TSS (Figure 12B/C). *IMD2* mRNA expression decreased 5-fold in *rtr1Δ* cells and when *RRP6* was deleted alone there was no statistically significant change (Figure 12D). However, when both *RTR1* and *RRP6* were deleted, *IMD2* mRNA levels did decrease relative to WT, but not quite to the same extent as with the *RTR1* deletion alone. In order to analyze the full-length transcripts that were produced in the cell, I performed Northern blots. The Northern blot analysis of *IMD2*'s coding region revealed similar results as the RNA-seq data at the coding region (Figure 12F). The data also confirmed that a mutant version of another serine-5 phosphatase, Ssu72-tov, displayed termination defects, which resulted in elevated expression levels of IMD2 [350]). These data suggest that Rtr1 may be acting as a negative regulator of NNS termination, compared to the Ssu72 and Rrp6 mutants which seem to decrease the NNS termination pathway efficiency.

5. Rtr1 Promotes Elongation of ASTs

The loss of Rrp6 in the *rtr1Δ* strain did not lead to full recovery of *IMD2* RNA levels as compared to the loss of Rrp6 in WT which had no effect. This suggests that Rtr1 may promote elongation of ncRNA in WT cells, and the change in RNA expression due to the loss of Rtr1 is not a change in stability of RNA but is an effect of Rtr1 attenuating or decreasing the efficiency of the NNS termination pathway. In order to confirm this hypothesis, I analyzed and compared global RNA-seq of both the *rrp6Δ and rtr1Δ rrp6Δ* transcriptomes relative to WT (Figure 13). When analyzing 104 ASTs that were significantly downregulated in *rtr1Δ* cells, I found that most of the Rtr1 regulated ASTs were dependent on the RNA exosome for the decreased expression of the ASTs (Figure 14). It is possible that the decrease in expression of the ASTs in the cells missing Rtr1 is related to increased premature ncRNA termination through the NNS pathway and requires the RNA exosome for degradation [383]. Previous research has shown that RNAPII elongation of ncRNA including ASTs can regulate the RNA expression of protein coding genes in what is called transcription interference [246]. Thus, it is possible that the increase in ncRNA elongation stimulated by Rtr1 in WT cells may be changing the global transcriptome.

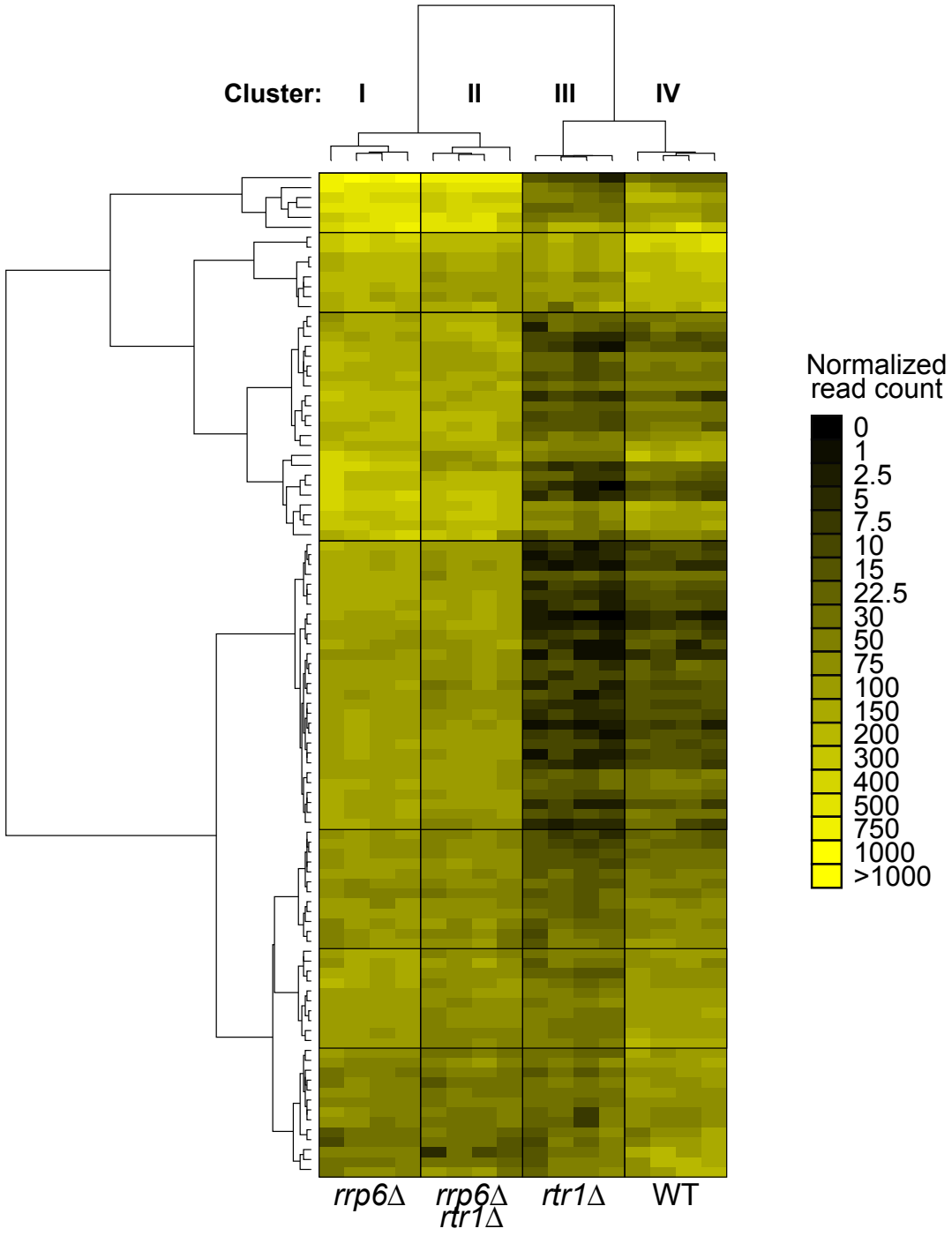

Cluster: I II III IV
rrp6Δ
rrp6Δ rtr1Δ
rtr1Δ
WT
Normalized read count
0
1
2.5
5
7.5
10
15
22.5
30
50
75
100
150
200
300
400
500
750
1000
>1000

Figure 14: WT, *rtr1Δ and or rrp6Δ* strains differentially expressed genes. Heat map of transcripts that were differentially expressed in *rtr1Δ, rrp6Δ,* and *rtr1Δ rrp6Δ* strains relative to WT cells. Scale at the right displays read count of each gene. Unsupervised hierarchical clustering analysis was performed using normalized read count values for each biological replicate and genotype (n=16). Each genotype shows all the biological replicates clustered together illustrating a high degree of reproducibility between replicates with WT values groups as Cluster IV, *rtr1Δ* values grouped as Cluster III, *rtr1Δ rrp6Δ* grouped as Cluster II, and *rrp6Δ* grouped as Cluster I.

Past studies have shown that transcription of the mRNA gene *YKL151C* is regulated by an antisense transcript, *YKL151C AS* , which is terminated by the NNS termination pathway (Creamer Corden 2011). This transcription interference of *YKL151C* by its antisense gene is modeled in Figure 15A. To confirm the RNA-seq results and to analyze the full-length RNA of the *YKL151C / YKL151C AS* genes, I performed Northern blot analysis using single-stranded RNA probes. The data show that the *YKL151C* sense transcripts are up-regulated in the *rtr1Δ* cells, a result possibly caused by *YKL151C AS* transcript downregulation (Figure 15B). The Northern data show an opposite effect in the *RRP6,* and *rrp6Δ/rtr1Δ* cells, which could be caused by decreased NNS termination leading to higher expression of sense *YKL151C* (Figure 15). In addition, the cells with the termination deficient Ssu72 mutation had decreased expression of sense *YKL151C* while its AST showed an increase in expression (Figure 15B). Finally, the RNA-seq analysis verified that *YKL151C* sense is upregulated in *rtr1Δ* cells while the *AS YKL151C* expression is decreased (Figure 15C/D).

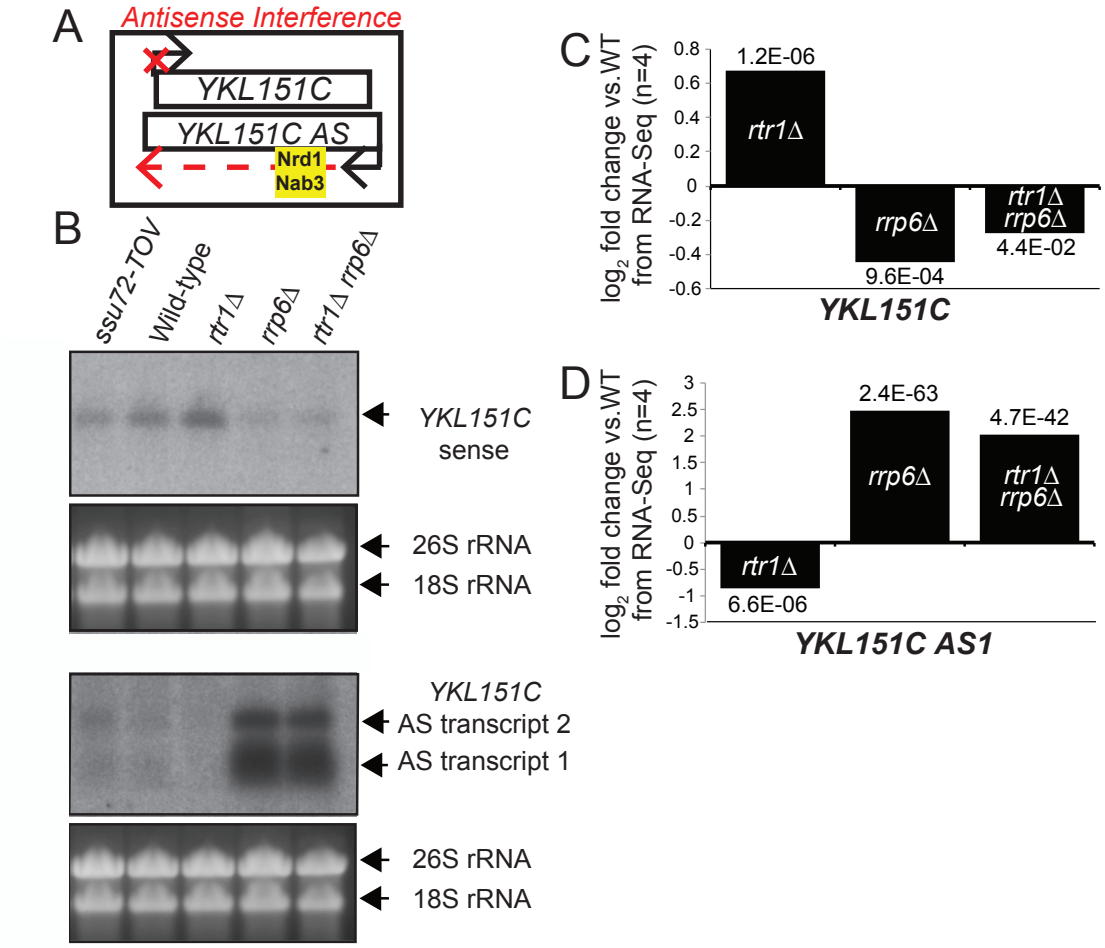


Figure 15: Analysis of YKL151C sense and antisense expression in WT and *rtr1Δ* cells. (A) Schematic representation of *YKL151C* sense and antisense expression. (B) Northern blot analysis of *YKL151C* sense and antisense transcripts using single-stranded RNA probes. (C) and (D) Summary of RNA-Seq expression analysis of *YKL151C* sense and antisense transcript expression, FDR values are displayed above/below bar.

Ssu72-L84F Mutant Results in Change in Transcription Termination Mechanism

Ssu72 is an essential gene so that previous studies have focused on Ssu72 mutations causing diminished function [296]. The first *SSU72* mutant described consisted of a duplication of amino acids 9-18 together with another mutation, which conferred a heat sensitive phenotype and shifted transcription start sites of select genes [296]. In the same study, it was discovered that Ssu72-C15S was lethal, and C15 was later confirmed to be a residue necessary for dephosphorylation activity [276, 296]. Relative to transcription termination, many studies have found that loss of function mutations or the depletion of Ssu72 leads to increased gene expression or modest read-through of termination sites [184, 185, 276, 277, 388, 389]. In this study, I focus on an Ssu72-L84F mutant yeast strain in which RNAPII has been shown to transcribe or read-through a termination signal upstream of the *IMD2* ORF (Figure 16) [350]. My goal was to investigate the effect of the Ssu72-L84F mutant on termination sites by analyzing RNAPII localization globally, through high resolution ChIP-exo, and by measuring transcript length and quantity through RNA sequencing and Northern blot analysis. My results show that the L84 residue of Ssu72 is important in regulating phosphatase activity and the mutation of L84 to F confers a gain of function that increases the phosphatase activity of Ssu72. In addition, the Ssu72-L84F mutation leads to increased RNAPII occupancy downstream of transcription termination sites as well as increased RNA expression across the genome.

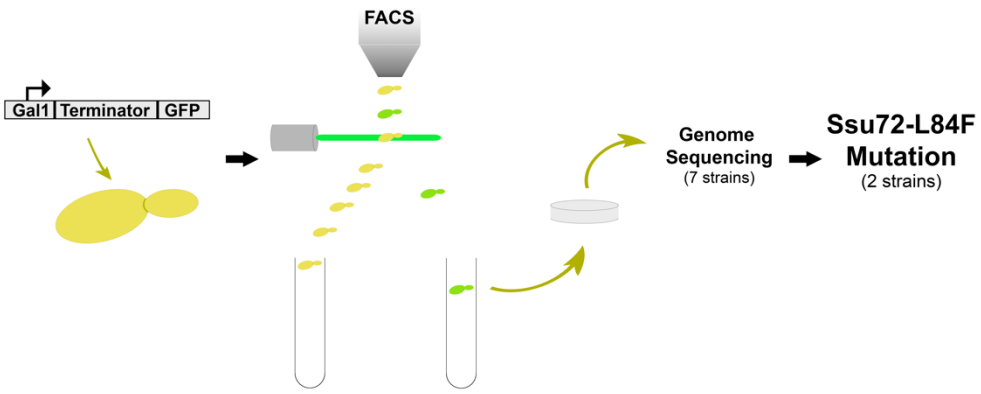


Figure 16: Spontaneous mutant Ssu72-L84F can bypass an intergenic terminator. Workflow of spontaneous mutant in Ssu72 possibly conferring termination over-ride [350].

1. Ssu72-L84F Mutation Confers Gain of Function Phenotype

To investigate if the L84F mutation affected activity of the Ssu72 phosphatase, a Western blot was performed on yeast whole cell lysate with antibodies specific to several forms of the phosphorylated CTD as well as an antibody specific to the CTD itself. WT BY4742 and Ssu72-L84F strains were used, both had a 3XFLAG tag on the Rpb3 subunit of the RNAPII complex to initially measure CTD phosphorylation. Initial results in two independently isolated clones found a decrease in signal in the serine-5 phosphorylated CTD as well as the serine-7 phosphorylated form of the CTD relative to the WT (Figure 17A). This result suggests that Ssu72-L84F is a hyperactive mutant, which is surprising, since previous Ssu72 mutants that caused read-through were loss of function mutants [184, 185, 276, 277, 388, 389]. In addition, the RNAPII subunit Rpb3 levels were unchanged in the mutant strain compared to WT, and the cytosolic protein Pgk1 show consistent loading across all of the samples. Taken together this suggests that the L84F mutation in Ssu72 decreases serine 5 and serine 7 phosphorylation of the RNAPII CTD. To further confirm these results, yeast strains were generated expressing Rpb3-3XFLAG in both WT and Ssu72-L84F strains with a 3C cleavage site inserted within the linker region of the CTD, before the repeating heptad sequence begins. RNAPII was immunocaptured through the Rpb3-3X FLAG tag on anti-FLAG antibody beads and while immobilized the CTD was released through 3C protease treatment and subsequently collected with the flow through material. CTD phosphorylation levels were measured using antibodies specific to Ser2-P CTD, Ser5-P CTD, Ser7-P CTD. Total CTD was used as a loading control and detected using a commercially raised antibody directed against the N-terminal portion of the CTD fragment that is released following 3C

protease cleavage in the strains. Surprisingly, phosphorylation was reduced in the Ssu72-L84F mutant compared to the WT at all three serine residues suggesting that Ssu72-L84F is a hyperactive variant that has lost specificity for serine 5 and serine 7 alone (Figure 17B). Additionally, the CTD migrated faster through the gel in the Ssu72-L84F mutant compared to the WT, suggesting an overall decrease in the bulk phosphorylation status of the CTD (Figure 17B). This change in CTD dephosphorylation activity could occur as a consequence of changes in CTD binding affinity and/or a change in phosphatase activity. In order to look at the specific activity of the mutant phosphatase, a phosphatase activity assay was performed with different mutants of Ssu72, at 2 μM (concentration determined by Bradford protein assay), including the catalytically inactive C15S, the WT, the Ssu72-L84F mutant, and the L84A mutant using a small molecule substrate. Using 10 μM DiFMUP (6,8-Difluoro-4-Methylumbelliferyl Phosphate)) as a substrate, the Ssu72-L84F mutant has significantly increased phosphatase activity relative to WT and the Ssu72-L84A mutant had even greater phosphatase activity than the Ssu72-L84F mutant (Figure 17C).

2. Global Changes in the Transcriptome

Ssu72 is an essential gene that plays a role in many transcriptional mechanisms and thus I hypothesized that the Ssu72-L84F mutant, which disrupts the termination mechanism of the *IMD2* terminator, would have a large effect on the transcriptome as a whole [350]. RNA-seq was performed to investigate any changes in the transcriptome in the Ssu72-L84F mutant compared to WT cells.

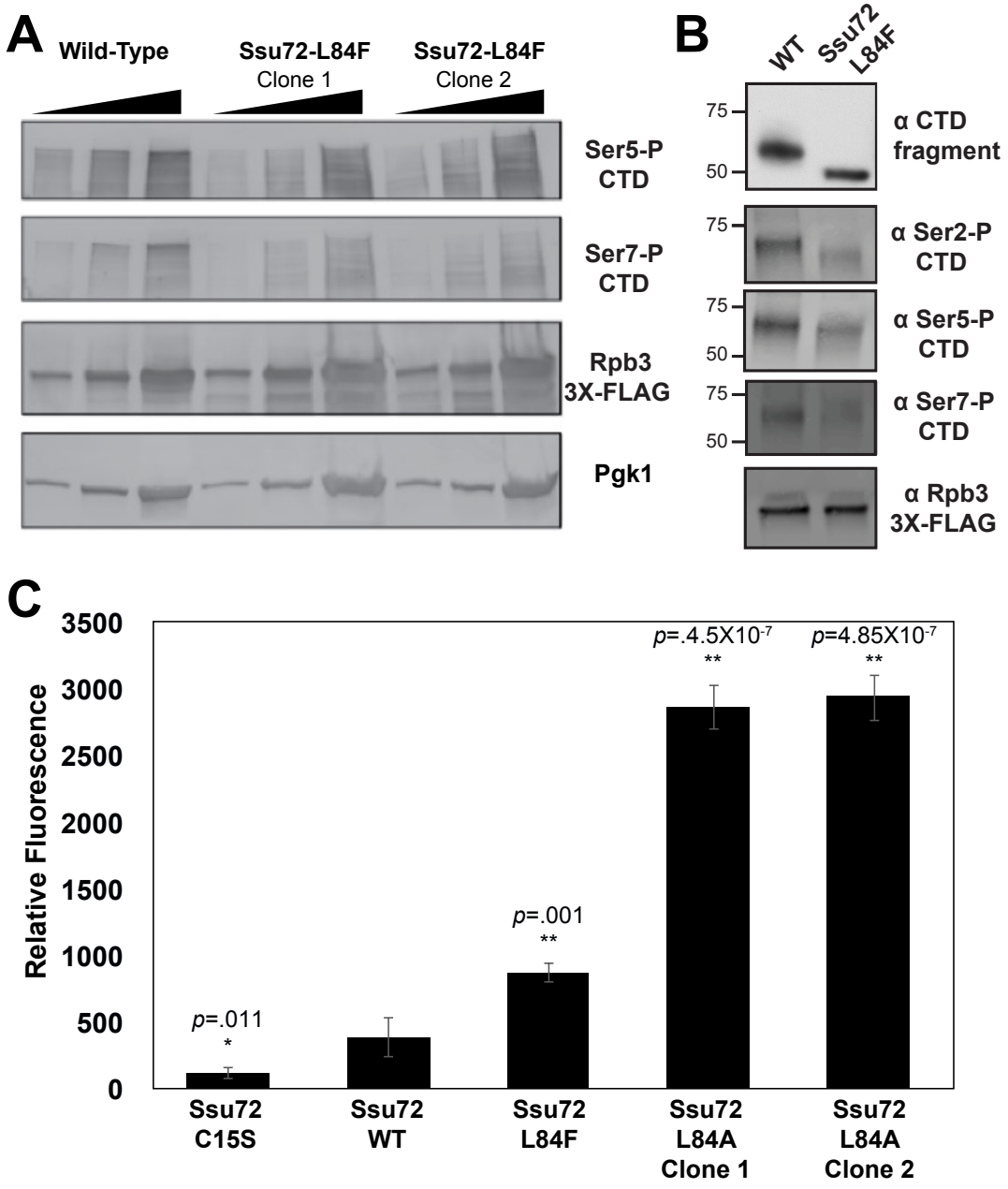


Figure 17: L84F mutation in Ssu72 affects phosphatase activity. (A) Western blot of whole cell lysate in WT and Ssu72-L84F strains. Antibodies specific for serine 5 phosphorylated CTD, serine 7 phosphorylated CTD, and loading controls RNAPII(Rpb3-FLAG) and PGK1. (B) Western blot on CTD from WT and Ssu72-L84F strains. Antibodies specific for serine 5 phosphorylated CTD, serine 7 phosphorylated CTD, serine 2 phosphorylated CTD, the CTD fragment itself and control RNAPII(Rpb3-FLAG). (C) Phosphatase activity assay using substrate DifMUP. Relative fluorescence was measured after 120 minutes at 450nm. C15S renders phosphatase inactive and abolishes cell viability. L84A and independently transformed L84A-2 mutants were created in the Mosley lab. Phosphatase activity assay using purified GST protein yielded results similar to C15S (data not shown)

To control for a large change in the transcriptome that could occur as a consequence of disruption of a central basal transcription regulator, I utilized a spike-in control (ERCC, Invitrogen) for a more accurate normalization of the RNA expression data. Considering that Ssu72 is involved in the CPF pathway and has previously been implicated in the generation of novel noncoding RNAs (ncRNAs), total RNA sequencing was performed, and a number of noncoding RNA transcripts annotations were included in the alignment steps. In brief, many regions of DNA previously thought to be transcriptionally inactive have been observed to be undergoing pervasive transcription [12, 156-158]. In addition, there is evidence that a major source of pervasive transcription occurs at bi-directional promoters and at nucleosome depleted regions (NFRs), resulting in a plethora of ncRNAs such as: stable unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs), Nrd1 unterminated transcripts (NUTs) and Ssu72 restricted transcripts (SRTs) [12, 13, 159, 160]. The function and purpose of many of these emerging ncRNA classes are yet to be completely understood. Most small ncRNAs, including sno/snRNAs are terminated via the Nrd1-Nab3-Sen1 (NNS) termination pathway in *Saccharomyces cerevisiae* which generally leads to non-coding transcripts being either degraded or processed [12, 130, 137, 147, 148, 159, 165, 237, 244, 250].

Our results combined RNA-seq data from four biological replicates and compared 11,296 annotations (including 5,794 ORF-Ts, 2,640 AST's, 1,157 NUTs, 711 SUTs, 434 CUTs, 214 SRTs, 80 snRNAs and 78 snRNA extended transcripts snRNA-ETs) (Figure 18A). Overall, there was a distinct trend of RNA overexpression genome-wide in the Ssu72-L84F strain relative to the WT (Figure 18A). A total of 3,232 transcripts, out of 11,296, were overexpressed and 755 transcripts had decreased expression (at an FDR less

than or equal to 0.1) which means that overall there was six fold more genes that were overexpressed. A number of mRNAs were significantly affected by the L84F mutation as 15% of the 5,794 annotations analyzed were differentially expressed (FDR < 0.1 Figure 18B). Non-coding RNAs were strongly affected by the Ssu72-L84F mutation, differential expression for each transcript class specifically was; snRNA-ET 71%, CUTs 60%, SRTs 58%, NUTs 56%, SUTs 45%, snRNAs 40%, and ASTs 34% (Figure 18B). Surprisingly, the differential expression profile for Ssu72-regulated transcripts (SRTs) in the Ssu72-L84F mutant was significantly different than that of another Ssu72 mutant (Ssu72-R129A), which also affected ncRNA transcription, indicating that each mutation acts in a distinct manner (Figure 19A). Of note, many of the Ssu72-R129A regulated SRTs were upregulated in *rrp6Δ* cells (n=105) and of those transcripts there was a high-degree of overlap with Ssu72-L84F (Figure 19B). These data suggest that the Rrp6-containing nuclear exosome may be involved in the regulation of many of the SRTs that are upregulated in Ssu72-L84F cells. This is consistent with previously described increases in SRT detection in Ssu72-R129A *rrp6Δ* double mutant cells [13]. Although Rtr1 and Ssu72 have both been characterized as CTD phosphatases, it is of note that *rtr1Δ* cells, only a small number of SRTs were up-regulated (n=7, data not shown) and instead more frequently had decreases in SRT expression (n=24, Figure 19B).

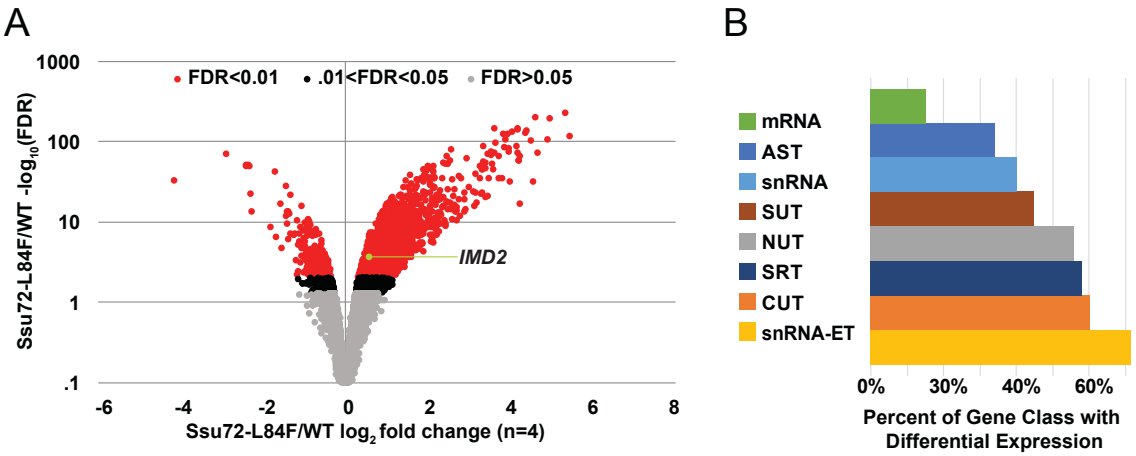


Figure 18: Differential expression analysis of whole transcriptome sequencing performed with ERCC spike-in controls for normalization. (A) Volcano plot illustrating Ssu72-L84F differential expression [-$\log_{10}$ FDR (y-axis) and fold-change (x axis)]. (B) Overview of transcript categories with varying degrees of differential expression changes in Ssu72-L84F. Extended transcripts (ETs) were observed for a high percentage of sn/snoRNA transcripts.

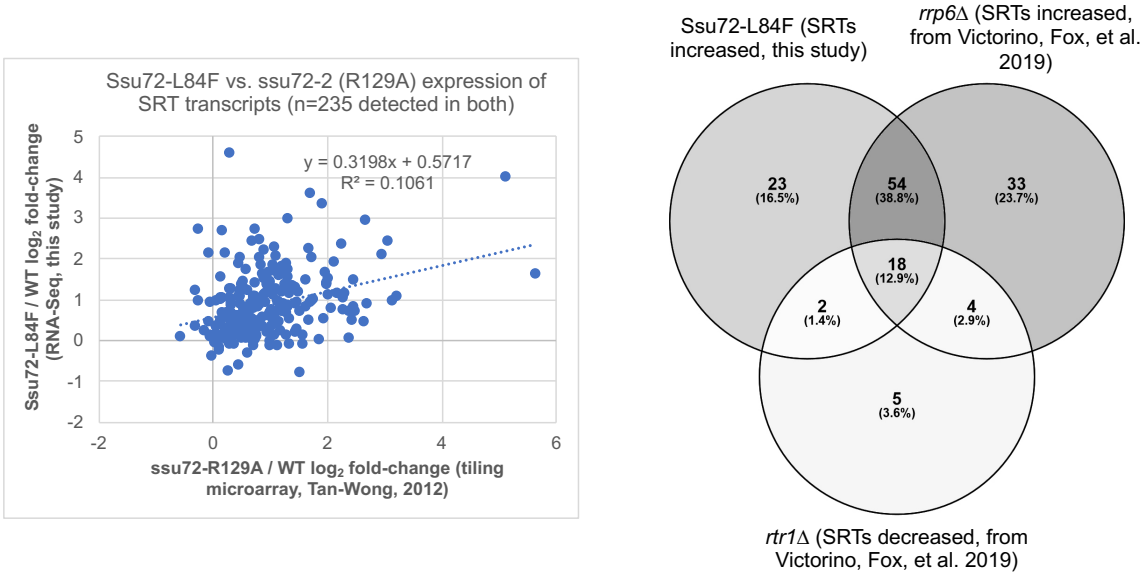


Figure 19: Differential expression analysis comparison of Ssu72-L84F strain and Ssu72-R129A strain. (A) Plot comparing the differential expression of annotated SRTs in the *ssu72*-TOV strain and *ssu72-2* strain. Gene expression of all annotated SRTs of the different strains were divided by WT expression and then the ratio of the fold change was converted to log of 2. Values for each strain were plotted relative to each other Ssu72-L84F on the y-axis and Ssu72-R129A on the x-axis. (B) Venn diagram comparing the different trains and the SRTs.

Finally, snRNA-extended transcripts (ETs, annotations of snRNAs that are downstream of the snRNA termination points, [248]) showed the largest percentage of differential expression for any gene category at 71% when compared to the 40% of snRNA transcripts, which provides evidence that the Ssu72-L84F mutant is causing read-through termination defects in these genes (Figure 18B). *rtr1Δ* cells showed the opposite phenotype to Ssu72-L84F cells for extended transcripts with more frequent decreases in snRNA ETs.

To gain a better understanding of how individual genes were affected, read counts per nucleotide at different genomic location which contained an snRNA gene were analyzed. Additionally, differential expression analysis revealed some downstream effect of snRNA gene readthrough because 85% of sn/snoRNA genes have an mRNA gene at least 500 nucleotides downstream of the snRNA termination site and in some cases the downstream gene was upregulated likely as a consequence of transcription readthrough. In the Ssu72-L84F mutant strain, *snR82* displayed a slight decrease in expression throughout the body of the gene, but at the 3' end near the termination site there was significantly more expression relative to the WT (Figure 20A). This trend continued through the mRNA gene *USE1* downstream of the snoRNA gene snR82. In order to verify that the increased RNA detected at USE1 was indeed a consequence of read-through of the snoRNA terminator, Northern blot analysis was done. I used a radioactive probe specific to the transcript from within the body of the *snR82* gene. The Ssu72-L84F mutant produced a transcript that in addition to the normal length of the snR82 transcript (268 bp) was ~1,800 nucleotides long, which is about the length from the beginning of the *snR82* gene to the end of the *USE1* gene (Figure 20B).

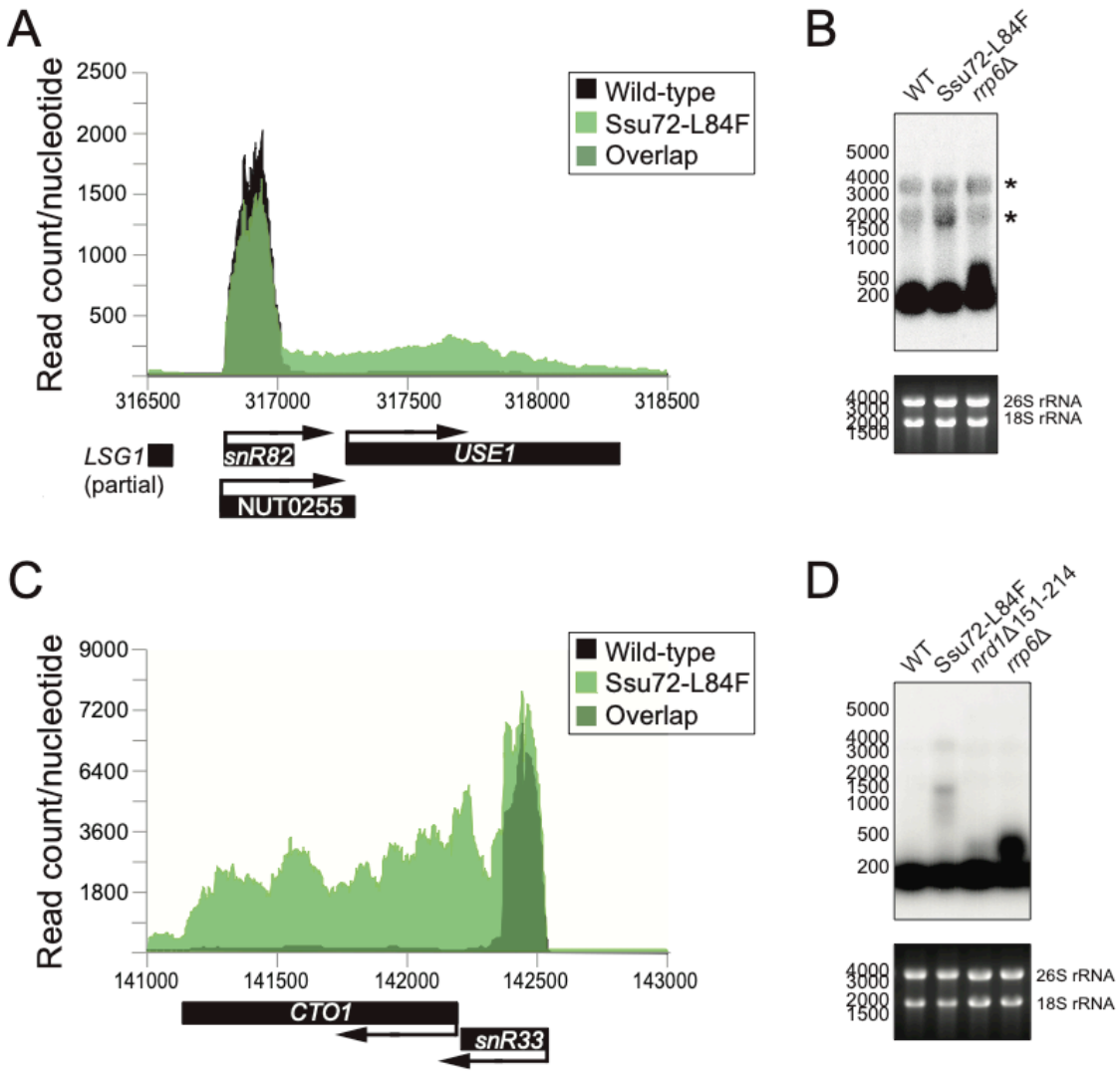


Figure 20: RNA-seq reads and Northern Blots of Ssu72-L84F at snR33 and snR82. (A)&(C) RNA sequencing reads at two regions of the genome showing the effects of the Ssu72-L84F mutant on gene expression at region around snR82 and snR33. Ssu72-L84F has increased gene expression downstream of the snRNA termination sites. (B)&(D) Northern blots showcasing the increased length of the ncRNA due to read-through. Northern blot with oligonucleotide specific to snR71. The Ssu72-L84F mutant has an extended snR71 transcript. Stars denote ribosomal RNA. Arrow denotes read-through transcript.

The WT strain only expressed transcripts of a single length of ~180 nucleotides. In this Northern I also analyzed an *rrp6Δ* strain which displayed two groups of transcripts, the normal length snRNA and a slightly longer transcript (~500 nucleotides) which represents the unprocessed snR82 transcripts. The Ssu72-L84F strain did not have the extra group of transcripts created in the *rrp6Δ* strain indicating that the effect of Ssu72-L84F is not solely occurring through changes in the action of the nuclear exosome. The snoRNA *snR33* had more RNA expression throughout the gene body in the Ssu72-L84F cells relative to the WT, but at the 3' end and downstream of the termination site there was a pronounced increase in RNA production across the *CTO1* gene (Figure 20C). The Northern blot of the snR33 RNA populations displayed a heterogeneous mix of transcript lengths in the Ssu72-L84F strain (Figure 20D). Along with the normal length snR33 RNA, there was a distinct group of transcripts that spanned the length of the *snR33* gene as well as the length of the *CTO1* gene (~1400 nt), as well as a smear of shorter transcripts (~800-1200 nt), and a smaller group of very long transcripts (~3500 nt). Again, the slightly longer group of snR33 transcripts (~300 nt), which are likely unprocessed forms form of the snoRNA, are observed in the *rrp6Δ* strain but are not present in the Ssu72-L84F strain. In the snR33 RNA Northern blot, an Nrd1 mutant with the deletion of the Nab3 interacting domain was also included. Other snoRNAs, such as snR71, have also been shown to produce a heterogeneous mixture of unterminated transcripts in the Ssu72-L84F strain (also known as Ssu72-TOV [248]) and other mutant backgrounds [381].

3. RNAPII Localization Analysis Indicates Termination Read-through

To analyze the effects that of mutant phosphatase Ssu72-L84F on RNAPII, I tracked the abundance of RNAPII throughout the genome using high-resolution ChIP-exo which we have previously employed to track RNAPII termination changes [248]. In this experiment, the third largest subunit of RNAPII, Rpb3, was tagged with 3XFLAG in the Ssu72-L84F strain at its chromosomal locus. For comparison, the impact of Ssu72-L84F on RNAPII occupancy was investigated at the snoRNA genes for snR82 and snR33. RNAPII localization was consistent with the changes observed with snR82 RNA expression, with RNAPII localized throughout the snR82 gene in both WT and Ssu72-L84F cells. However only in Ssu72-L84F RNAPII appears to continue downstream of the snR82 termination site perhaps accumulating at the TSS of *USE1* but also continuing downstream into the *USE1* gene (Figure 21A). These data suggest that RNAPII reads through the snR82 terminator but is eventually terminated at the USE1 termination site, likely through a polyA-dependent termination mechanism. Additionally, it is possible that the area around the *USE1* TSS show RNAPII accumulation as a consequence of competition between RNAPII that is continuing to transcribe past the snR82 gene but is subsequently competing with the RNAPII that is trying to transcribe the *USE1* gene. Use1 is an essential protein so it is highly likely that a sufficient amount of Use1 mRNA is being produced to maintain its expression. RNAPII showed a similar trend on snR33 with somewhat lowered accumulation of RNAPII across the 5' end of the snR33 gene and increased RNAPII levels throughout the *CTO1* gene, especially at the TSS (Figure 21B).

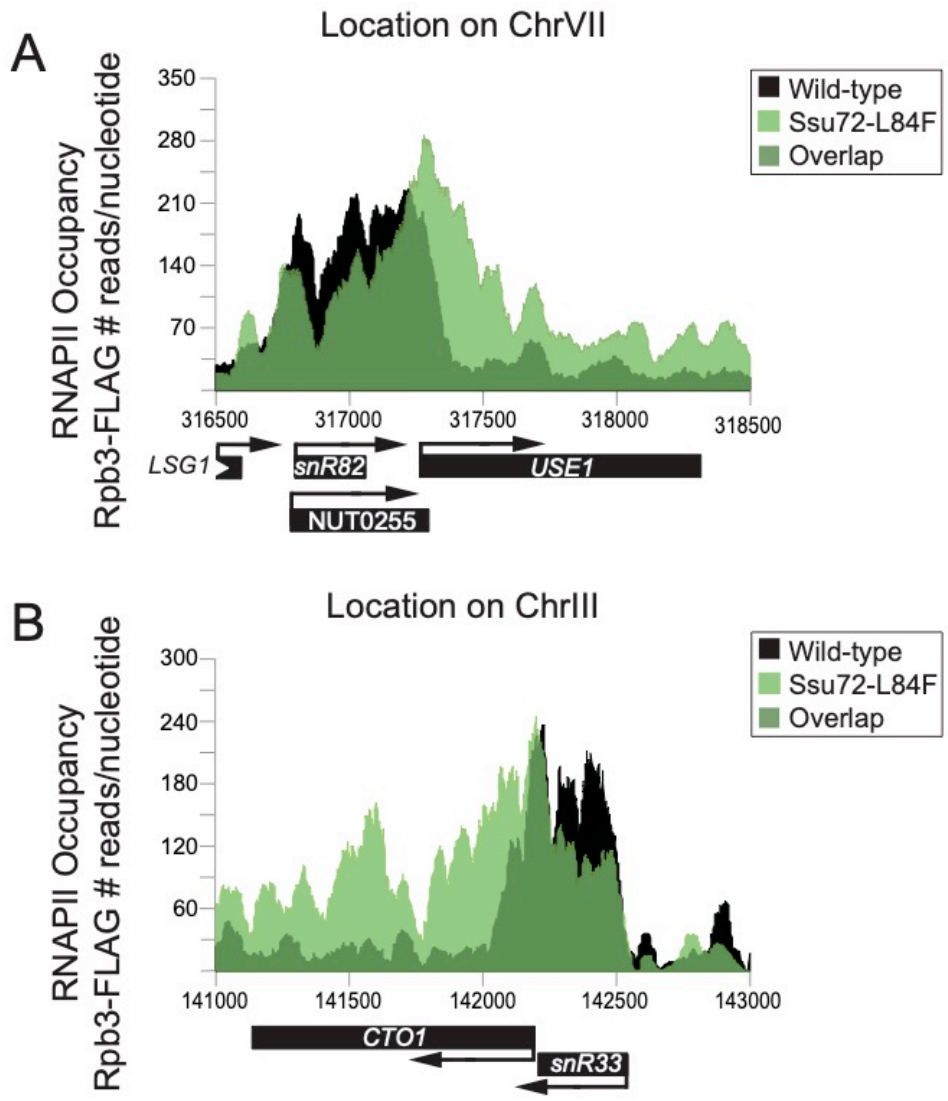


Figure 21: RNAPII occupancy of Ssu72-L84F at snR33 and snR82. (A)&(B) Chromatin immunoprecipitation and exonuclease analysis (ChIP-exo) of Rpb3-FLAG data at two regions of the genome showing the effects of the Ssu72-L84F mutant on the occupancy of RNAPII. Localization of RNAPII at region around snR82 and snR33 gene; Ssu72-L84F has a shifted RNAPII localization downstream of the snRNA termination sites and a reduced accumulation of RNAPII at the snRNA genes.

EdgeR was used to measure differential RNAPII localization and 594 gene annotations showed signs of differential RNAPII localization at an FDR $\leq$ 0.01 when comparing the Ssu72-L84F mutant to the WT with most of the annotations showing increased relative RNAPII levels in the Ssu72-L84F mutant, which is in agreement with the RNA-Seq dataset (compare Figure 18A and Figure 22). At an FDR $\leq$ 0.05 a total of 938 gene annotations showed changes in RNAPII occupancy. Increased RNAPII occupancy was measured at a number of snoRNA targets including snR37 as well as snR68 which has previously been found to have RNAPII readthrough into the *FLX1* mRNA as a consequence of Nab3 depletion using an anchors away approach (aqua dots with labels, Figure 22; [390]). A number of protein coding genes also showed increased RNAPII occupancy such as *IMD2*, which has previously been shown to have increased mRNA production in Ssu72 mutant cells [350]. *YDR042c*, a protein coding gene of unknown function, was also increased in Ssu72-L84F cells. *YDR042c* was initially identified as a protein coding gene that was expressed as a consequence of the *ssu72-ts69* mutation [391]. The sequence changes that occur in *ssu72-ts69* have not been determined, but it was found to have readthrough transcription at a number of snoRNAs similar to Ssu72-L84F. Differential RNAPII occupancy was also observed at genes that have not previously been shown to be regulated by Ssu72 such as Yvh1, which encodes a dual specificity protein phosphatase [392].

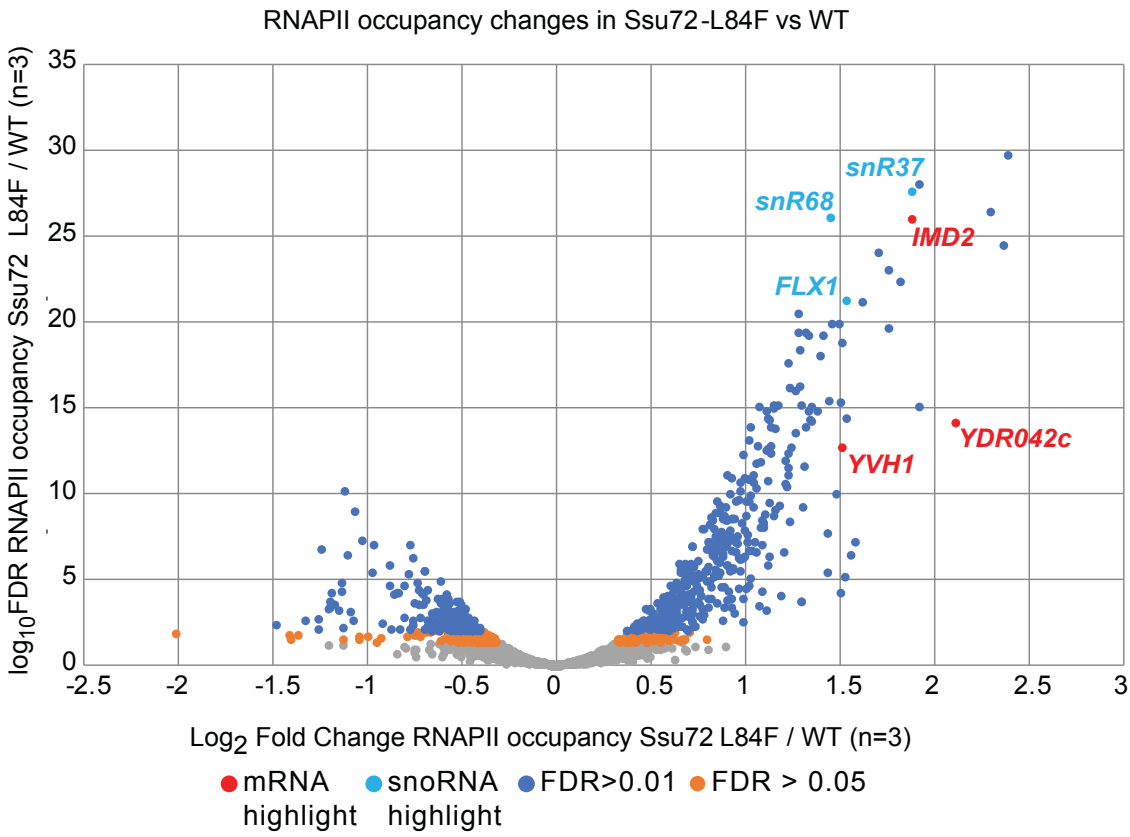


Figure 22: ChIP-exo differential RNAPII occupancy. Volcano plot illustrating Ssu72-L84F differential RNAPII occupancy at genes. Differential occupancy was calculated using the edgeR package but used ChIP-exo data instead of using RNA-seq data for the analysis. Legend displays the color-coded FDR values.

Through more in-depth characterization of the target genes altered by the Ssu72-L84F, there is potential to obtain unique insights into the precise mechanisms that underlie the requirement for Ssu72 in RNAPII transcription termination. However, these in-depth characterization experiments are outside the scope of this current work.

Global average RNAPII occupancy was plotted at sn/snoRNAs and mRNA genes in WT and Ssu72-L84F cells from ChIP-exo data (n=3 for each genotype). The plots for the snRNA genes were respective to the transcription start site (TSS) and termination end site (TES) and showed similar findings as previously shown single gene examples (Compare Figure 23 to Figure 21). The average RNAPII profile in the Ssu72-L84F strain had an extended peak downstream of the TES, and this trend continued ~600 nucleotides downstream of the TES. These data suggest that RNAPII transcription readthrough at sn/snoRNA terminators occurs at many of the 73 target genes in this class (Figure 23). However, for protein coding genes, more unique insights can be gleaned from the average gene analysis than are obvious from single gene analysis. At protein coding genes, RNAPII occupancy at the TES shows increased overall occupancy that extends ~100 nucleotides downstream but then returns to WT levels of RNAPII occupancy by ~300 nucleotides (Figure 24B). These data suggest that RNAPII termination at protein coding genes may be less efficient in Ssu72-L84F cells than WT leading to accumulation of RNAPII at termination sites (Figure 24B). Similar average RNAPII profiles were generated but aligned to the TSS rather than the TES.

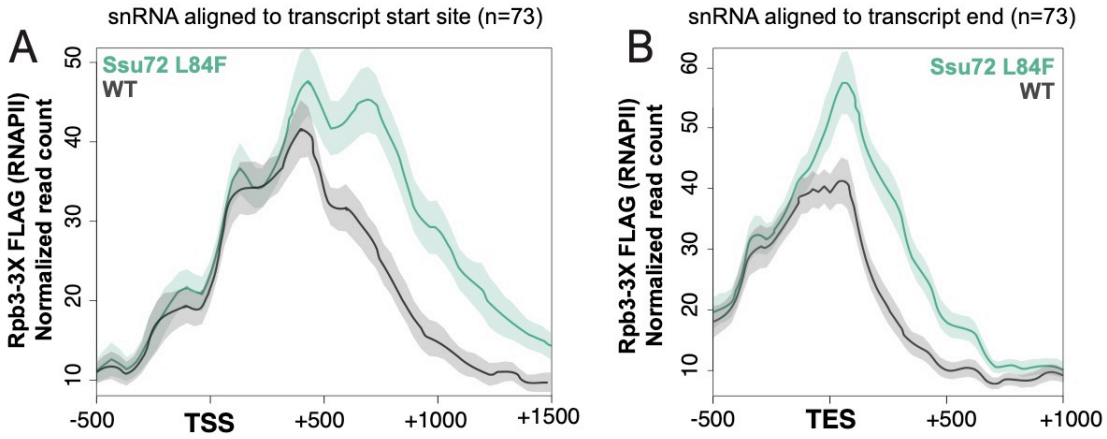


Figure 23: SNR Global ChIP-exo analysis using NGS.plot program. Graphs are centered around transcription end sites and transcription start sites of each gene. Read counts are averaged across genes in the annotation list and plotted. (A) TES of snRNA genes how a clear shift and increase of RNAPII localization. (B) TES of protein coding genes also show a less significant shift and increase of RNAPII localization.

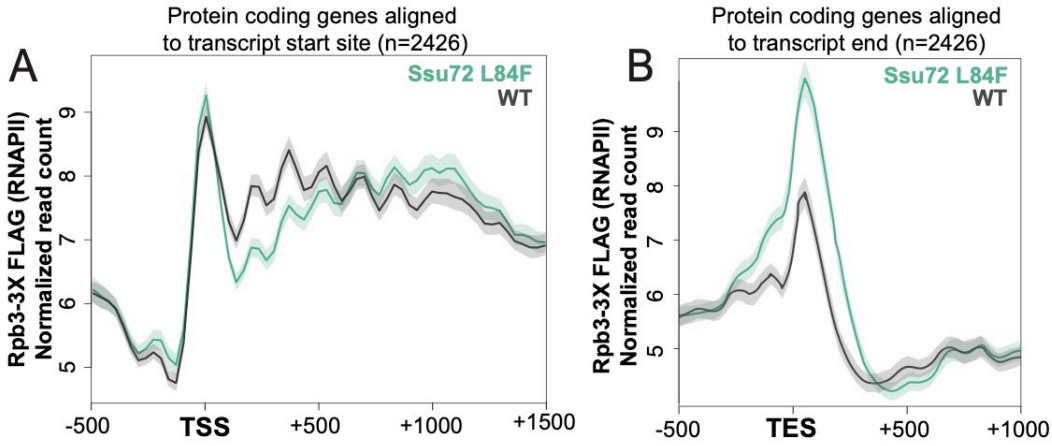


Figure 24: Protein coding Global ChIP-exo analysis using NGS.plot program. non-overlapping protein coding genes were analyzed. Graphs are centered around transcription end sites and transcription start sites of each gene. Read counts are averaged across genes in the annotation list and plotted. (A) TSS of snRNA genes show a shift and increase of RNAPII downstream of the TSS. (B) TSS of protein coding genes show a more interesting picture. Here there is a drop in RNAPII localization during early elongation and then an increase downstream of TSS site. This gives evidence that Ssu72 and the NNS pathway might be affecting protein coding genes and is not limited to ncRNA

While increased levels of RNAPII were still obvious towards the 3' end of RNAPII target genes, overall occupancy of RNAPII did not seem to be significantly altered at the TSS for either the snRNA genes or the protein coding genes (Figure 23A compared to 22A). Interestingly, protein coding genes did show significantly decreased RNAPII occupancy in the Ssu72-L84F mutant during early transcription elongation which recovers in late elongation as RNAPII nears the 3' end (Figure 24).

Early RNAPII elongation is associated with high levels of serine 5 phosphorylated RNAPII CTD [180, 393]. Nrd1 is one of the few proteins with a CTD interacting domain that strongly favors a serine 5 phosphorylated CTD during early elongation although serine 5 phosphorylation is also required for capping enzyme recruitment. I hypothesize that the Ssu72-L84F mutant is disrupting either the recruitment or stable association of the Nrd1 subunit of the NNS complex. This change in Nrd1-RNAPII interaction likely occurs as a consequence of the hyperactive phosphatase activity which leads to overall decreases in serine 5 phosphorylation in Ssu72-L84F cells (Figure 17). When comparing the global Nrd1 localization ChIP-exo data in WT cells to the Ssu72-L84F ChIP-exo data the peak of Nrd1 occupancy overlaps the region where RNAPII occupancy is reduced in the Ssu72-L84F mutant. I conclude that during normal elongation RNAPII progression is slowed by various molecular events, such as Nrd1 binding and RNAPII navigation around well positioned histones, thereby increasing the relative occupancy of RNAPII during early elongation. However, if Nrd1 binding is reduced as a consequence of hyperactive removal of serine 5 phosphorylation from the RNAPII CTD, relative RNAPII occupancy is decreased. These data provide evidence that

Nrd1 may participate in an early elongating checkpoint by slowing RNAPII perhaps to facilitate a mRNA 5' end capping checkpoint.

Ssu72-L84F Mass Spectrometry Analysis Reveals Sub-Complex Details of the CPF

Among the many pathways involved in transcription is the polyadenylation dependent transcription termination for which requires CPF and CF1A/B for termination and processing of mRNA [112, 117, 118, 237, 256-260]. The exact mechanism by which these complexes interact within and between each protein/RNA complex is still unknown and elusive. One observation is that there exist subcomplexes within the CPF complex as first described by the Greenblatt group and recently confirmed using non-covalent electrospray ionization mass spectrometry by the Passmore/Robinson group [256, 284]. The Passmore group was able to identify a subunit of the CPF complex, Syc1, which is not part of to the entire complex but only a part of the APT [282]. They found that presence of the Syc1 protein promoted SNR gene expression compared to the whole of the CPF complex that normally targets mRNA genes at the 3' end.

In this study, I focused on a *Saccharomyces cerevisiae* protein complex CPF. I used the phosphatase mutant Ssu72-L84F that has increased phosphatase activity which results in global readthrough of SNR genes and results in differential protein-protein interactions with different members of the CPF complex, specifically with Swd2, Glc7 and Ref2 of the APT subcomplex. To study the differential PPIs during termination as a result of the Ssu72-L84F variant, I optimized a coupled affinity purification XL-MS method to identify if there are any differences in the architecture structure of the CPF and CF1A complexes. Although these complexes are not the most abundant in the cell, there are many members and interacting partners. In addition, I also performed XL-MS on

RNAPII as it is an abundant and tight complex in the cell that has numerous well documented protein protein interactions (PPI) within and outside of the complex. Of note, since there are a number of existing high-resolution crystal structures of the RNAPII complex, I can make comparisons between the existing structural data and my crosslinking data as a validation approach.

1. Ssu72-L84F Mutation Leads to Change in Protein-Protein Interaction Network

In order to look at how the Ssu72-L84F mutation might affect the protein-protein interaction network, I used affinity purification mass spectrometry. I put a FLAG tag on Pta1, which is thought to be a scaffold protein in the CPF complex and more specifically in the APT subcomplex, which stands for Associates with PTA1. This protein is best suited for this AP-MS study because crystal structures for Ssu72 and Pta1 are available, and the activity of Ssu72 is increased when Pta1 is present in vitro. After AP-MS, the samples were subjected to SDS-PAGE and silver stain to make sure my purifications are successful and clean of any visible contaminants such as human keratins (Figure 25). This figure shows one of the biological replicates for a WT strain and a Ssu72-L84F strain. The purifications were abundant and free of contaminants which can be analyzed more clearly in elution 3. I performed four successful purifications for each strain which were digested and then detected on a Q-Exactive H-FX mass spectrometer. The spectra detected on the Q-Exactive H-FX were analyzed on the Proteome Discoverer software (Thermo) to match the spectra to the proteins the daughter ions belonged to.

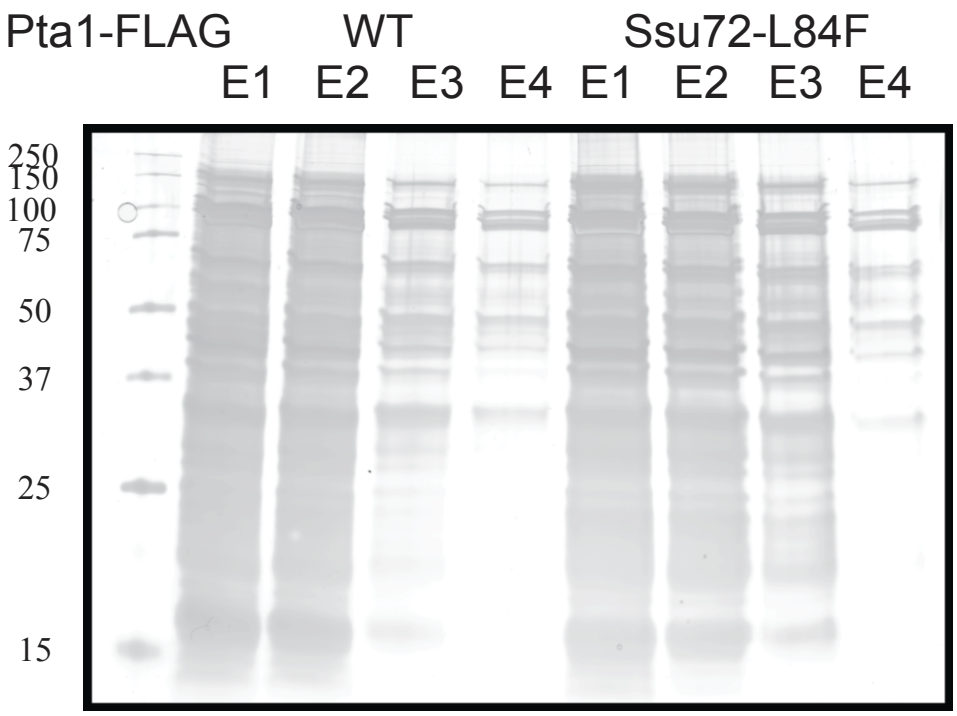


Figure 25: Affinity purification of termination complex subunits. Silver-stain of eluted proteins interacting with bait protein. Pta1 is FLAG tagged in WT and Ssu72-L84F strain, stain shows several subunits of the CPF complex. Elution of interaction proteins from the same sample are in order.

SAINT analysis was performed on three biological replicate purifications in order to identify the probability that the proteins detected were real interactions and not non-specific binding to the bait. The results for the all of the identified proteins were plotted by SAINT probability and fold change (Figure 26). Fold change B from the figure is calculated by comparing the total peptide spectral matches to the three mock purifications with the highest number of detected spectral matches. All of the CPF complex members had a SAINT probability of one which means that the proteins identified have the highest possible probability that they are indeed interacting partners. In addition there were many proteins involved in the transcription process that also co-purified of Pta1. All of the CPF complex proteins have a high ratio of sample to mock fold change values providing even more evidence that they are real interactions which is expected for proteins that are in complex with each other. Most of the proteins with the highest fold-change scores were also CPF complex members meaning that my purifications were relatively clean and that there are probably not too many proteins that interact with Pta1 as strongly as the members of the CPF complex.

In order to see if there is a difference in the protein protein interaction network due to the Ssu72-L84F mutation, I plotted a PPI map with color coded strings to represent a range of the SAINT probabilities and circles to represent the interacting proteins (Figure 27). In this network I focused on four different groups of proteins whose interactions I hypothesized could be affected by the Ssu72 mutation; CPF, CF1A, RNAPII, and NNS. The CPF group could not be compared with SAINT analysis due to the fact that all of the proteins had interaction scores of 1 and did not change with the mutation. This suggests that the Ssu72 L84F mutation does not cause major changes in

the association of Ssu72 with APT/CPF. RNAPII displayed significant changes in interactions due to the mutation in Ssu72. The Ssu72-L84F mutation increased the number of RNAPII subunits that interacted with Pta1, specifically Rpb1, Rpb3, Rpb9 and Rpb11, while it increased the probability score that Rpb2 had previously from 0.74-0.15 to >0.95,this suggests that the mutation is causing Pta1 to interact more with RNAPII.

This is consistent with the ChIP-exo findings in Figure 24 that show that RNAPII levels increase at the transcription end site (TES) which is the location at which the CPF complex should also be enriched. The CF1A complex and associated proteins did not show as much change as RNAPII. NNS complex proteins also displayed no significant change in the SAINT probability scores.

2. Ssu72-L84F Mutation Differential PPI Analysis Reveals Different Mechanistic Roles in CPF complex

Because the CPF was very abundant in the AP-MS study, all of the proteins in the complex had the highest SAINT probability score of 1. This makes it hard to see if there are any differences in the interactions within the complex. Therefore, I employed a method using the same SAINT analysis algorithm and altering it to analyze the differential PPI between two samples. SAINT analysis is usually performed by comparing a sample to a mock purification which gives the probability that an interaction is a true interaction. Instead, I used SAINT to compare two different samples to each other setting the WT as though it were a mock purification (Figure 28).

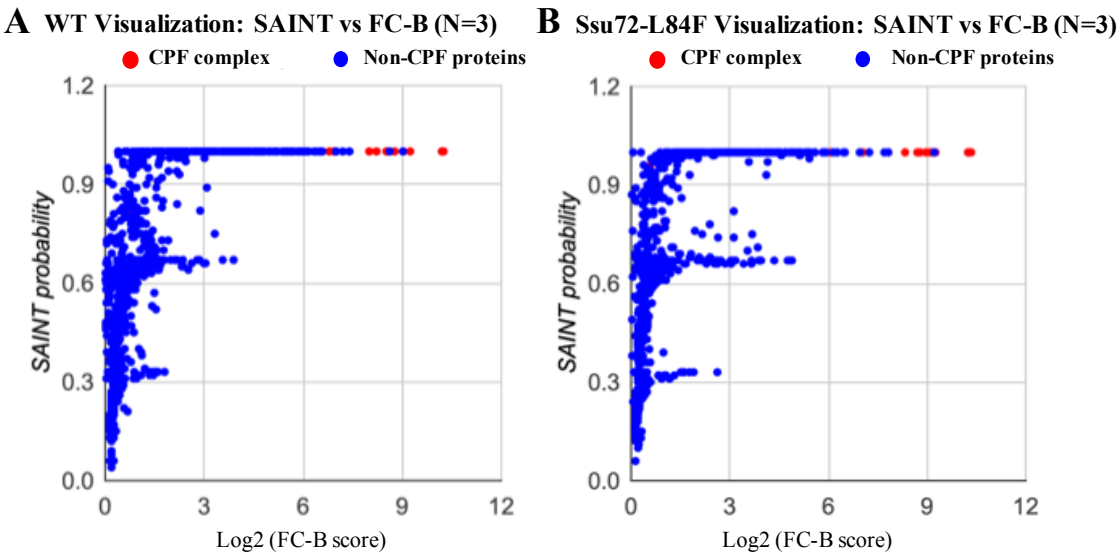


Figure 26: SAINT analysis of WT and Ssu72-L84F. (A) Fold Change of peptide spectral matches from WT plotted against SAINT probability (of a true interaction) (B) Saint probability vs. fold change for Ssu72-L84F

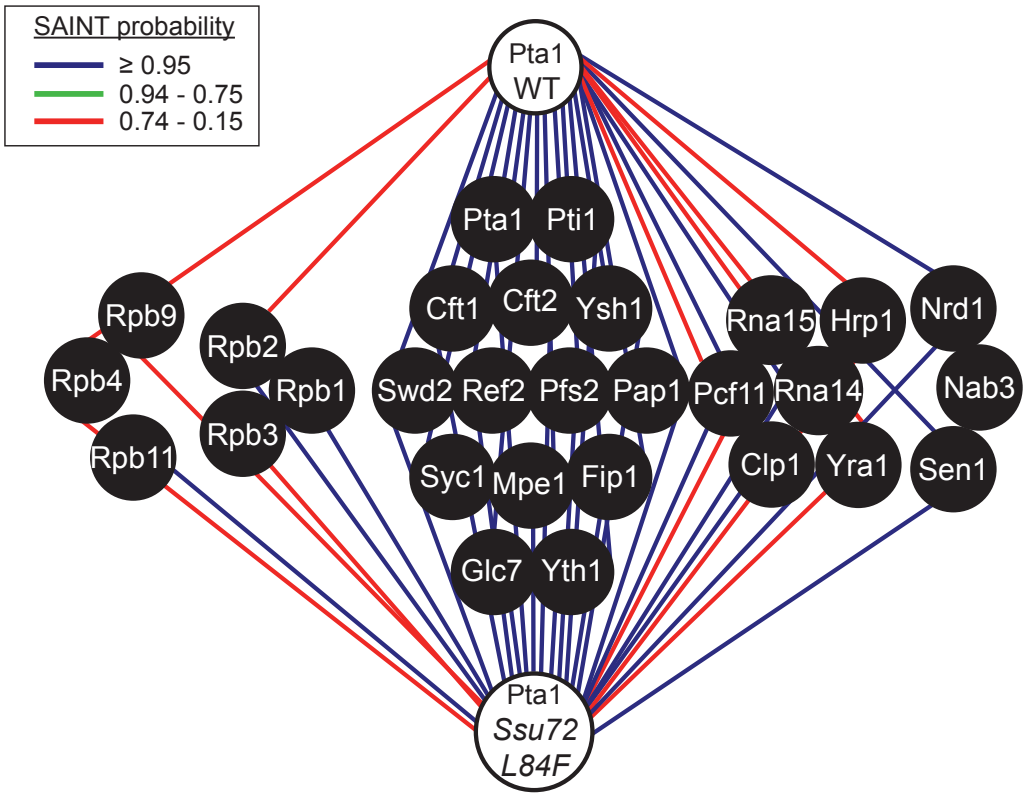


Figure 27: Purification of RNAPII termination complex. Pta1 is tagged with 3XFLAG sequence and purified using FLAG antibodies. Protein-protein interaction analysis obtained from SAINT analysis of LC-MS/MS raw data of purified termination factor complex. Colors of strings denote SAINT probability range.

Using this method, we calculated a statistical SAINT value for the probability that there is a difference between specific protein interactions with the bait in the Ssu72-L84F mutant compared to the WT. This analysis highlighted a set of four proteins, Glc7, Ref2, Swd2 and Ubx1, which had a high probability that they were differentially interacting with Pta1 relative to WT. This is interesting because all of these proteins are part of the APT complex except for Ubx1, a ubiquitin binding protein which is known to positively regulate Glc7 activity[394].In addition, these three CPF proteins have been found in a previous study to fragment together in ESI-mass spectrometry from the CPF complex relative to the other subunits[284]. This provides evidence that there may be another subcomplex within the APT subcomplex and that the Ssu72-L84F mutation is regulating the formation of this subcomplex. It is also possible that this subcomplex is regulating NNS termination in an independent mechanism from the CPF termination mechanism at the 3' end of mRNA genes. Other CPF complex proteins also had a relatively high probability of differential PPI, but the SAINT probability scores for these proteins were not high enough to confidently say that they were true differential interactions as a result of the Ssu72-L84F mutant. As expected, the Pta1 protein had a very low probability score since it cannot differentially interact with itself as it is the bait protein. Interestingly other transcription proteins had high probability scores such as NNS complex proteins Nrd1/Nab3/Sen1, RNAPII complex protein Rpb11and 3' end processing protein Xrn2.

3. Development of Protein Complex Cross-Linking of in Vivo Proteins

In order to better understand how proteins might differentially interact with each other at the residue level, I sought to develop a method which uses cross-linking mass

spectrometry with my affinity-purification mass spectrometry studies. This method utilizes a CID cleavable crosslinker called DSSO that can be identified in the mass spectrometer and in a package in Proteome discoverer called XlinkX. During the development of this method, 5 and 4 biological replicates were performed for Ssu72-L84F and WT cells. Of those that yielded cross-linking results, there were 35 unique cross-links per technical replicate (3 per biological replicate) (Figure 29). I was able to uncover several unique intra and inter crosslinks but was not able to recover intralinks from all of the different CPF complex proteins. In order to verify my experiment, I compared some of my crosslinks to a cryo-EM structure of Cft1 and Pfs2 (Figure 30). The distance between the nitrogen atom of the lysine side chain residue was measured for two intralinks and one interlink. One of the two lysine residues Cft1 lysine 211 was not included in my map since it had lower probability of being a true crosslink according to xlinkx software. This result is interesting considering that the distance, 15.2 Å, between nitrogen atoms was longer than the other two crosslinks and longer than the DSSO molecule. The other two crosslinks' nitrogen atoms were closer to each other and the distance between nitrogen atoms was 8.7 Å for the intralink and 11.7 Å for the interlink. This gives me confidence that my crosslinks more than likely are true crosslinks.

In addition to the CPF cross-linking experiments, I also carried out cross-linking experiments with the CF1A complex and RNAPII (Figure 31). The CF1A complex had very few cross-links while the RNAPII complex had several hundred unique cross-links. It is likely that larger protein complexes that are more abundant can yield better cross-linking results.

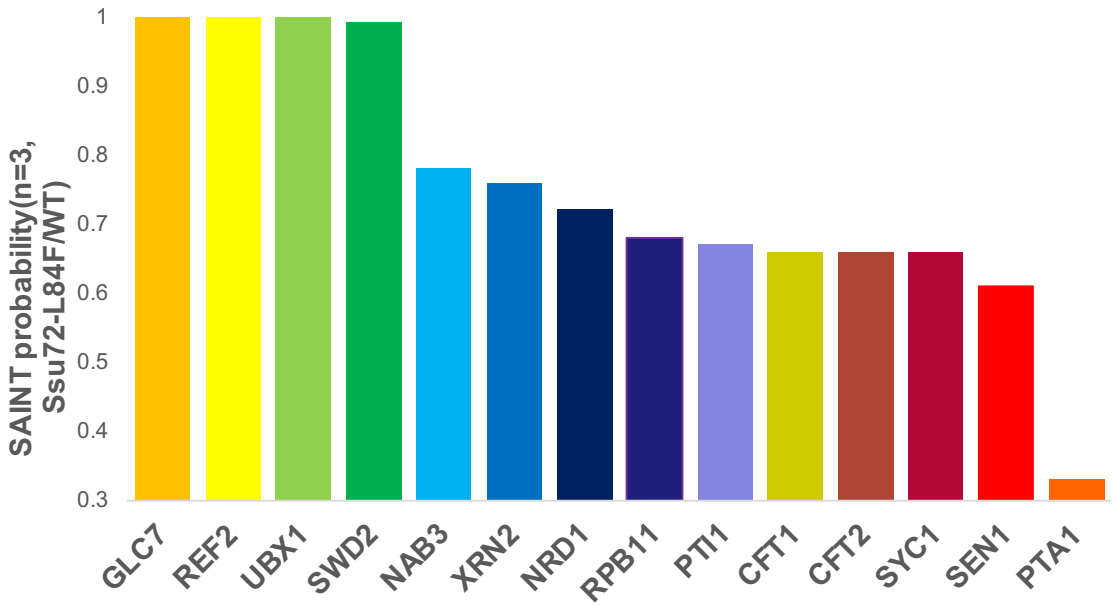


Figure 28: Differential protein-protein analysis in the Ssu72-L84F mutation using SAINT analysis algorithm. WT data was used as a mock control to statistically distinguish two different groups of data and identify real differential interactions between Ssu72-L84F and the WT.

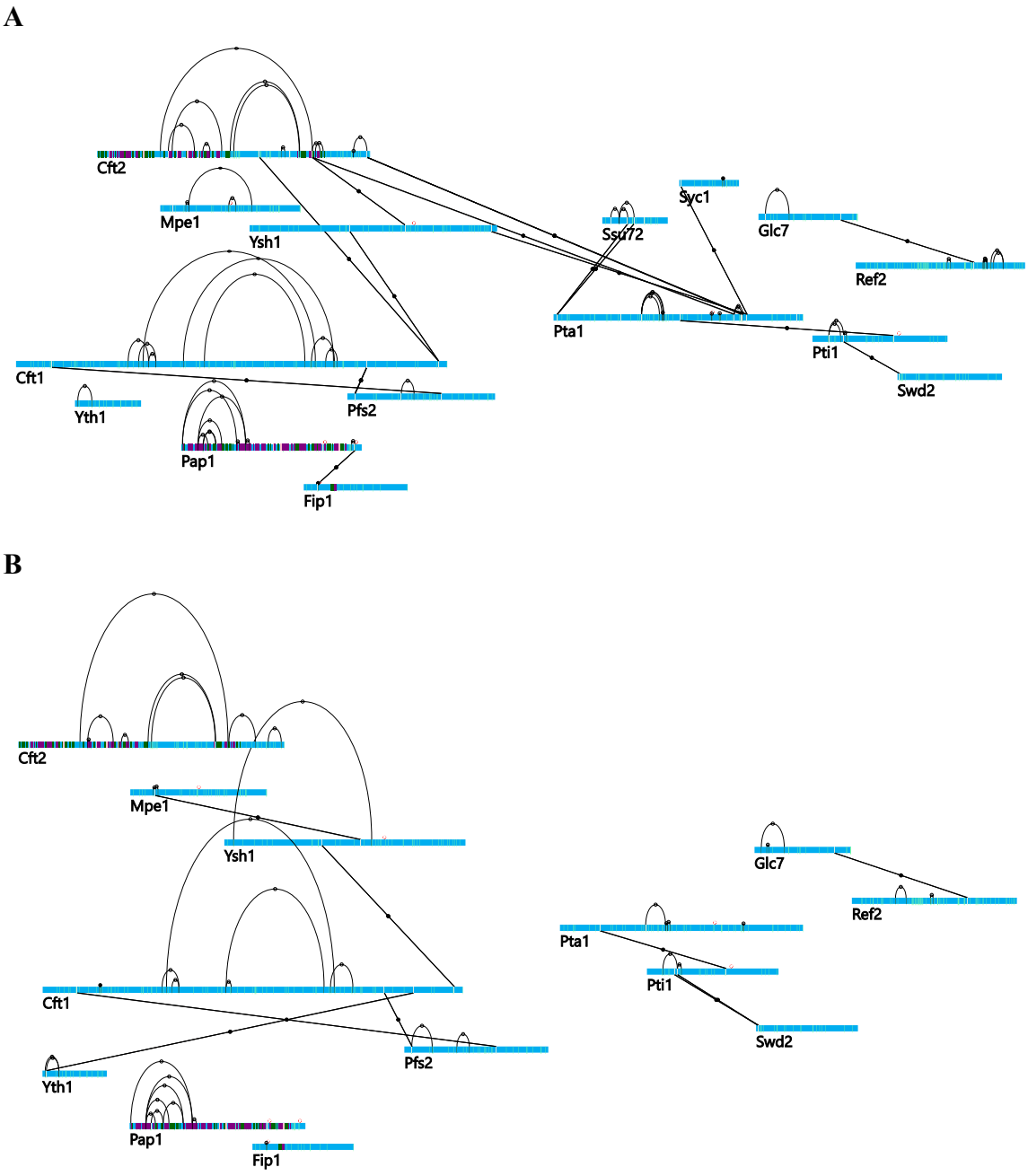


Figure 29: DSSO XL-MS analysis of CPF complexes by MS2-MS3 CID on an Orbitrap Fusion Lumos. (A) Ssu72-L84F CPF complex mapping intra-links and inter-links (B) WT CPF complex mapping intra-links and inter-links.

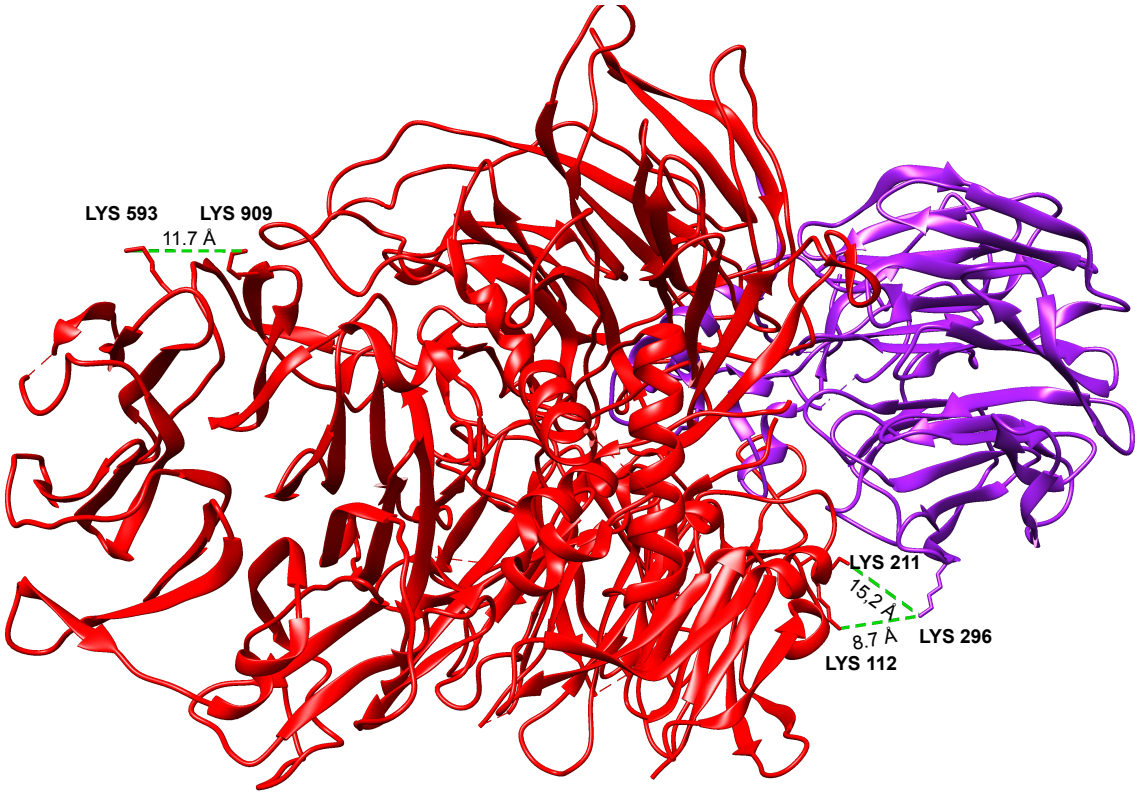
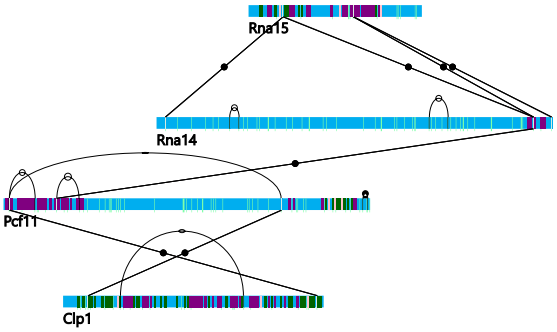


Figure 30: DSSO XL-MS verification with cryo-EM structure of Cft1 and Pfs2 of CPF complex. Detected crosslinks are verified by measuring predicted distances from the nitrogen atoms of the lysine side chains. Cft1 is in red and Pfs2 is in purple. DSSO spacer length is 10.1 Å. Original structure was determined at 3.5 Å resolution (PDB ID:6EOJ)[284]

A



B

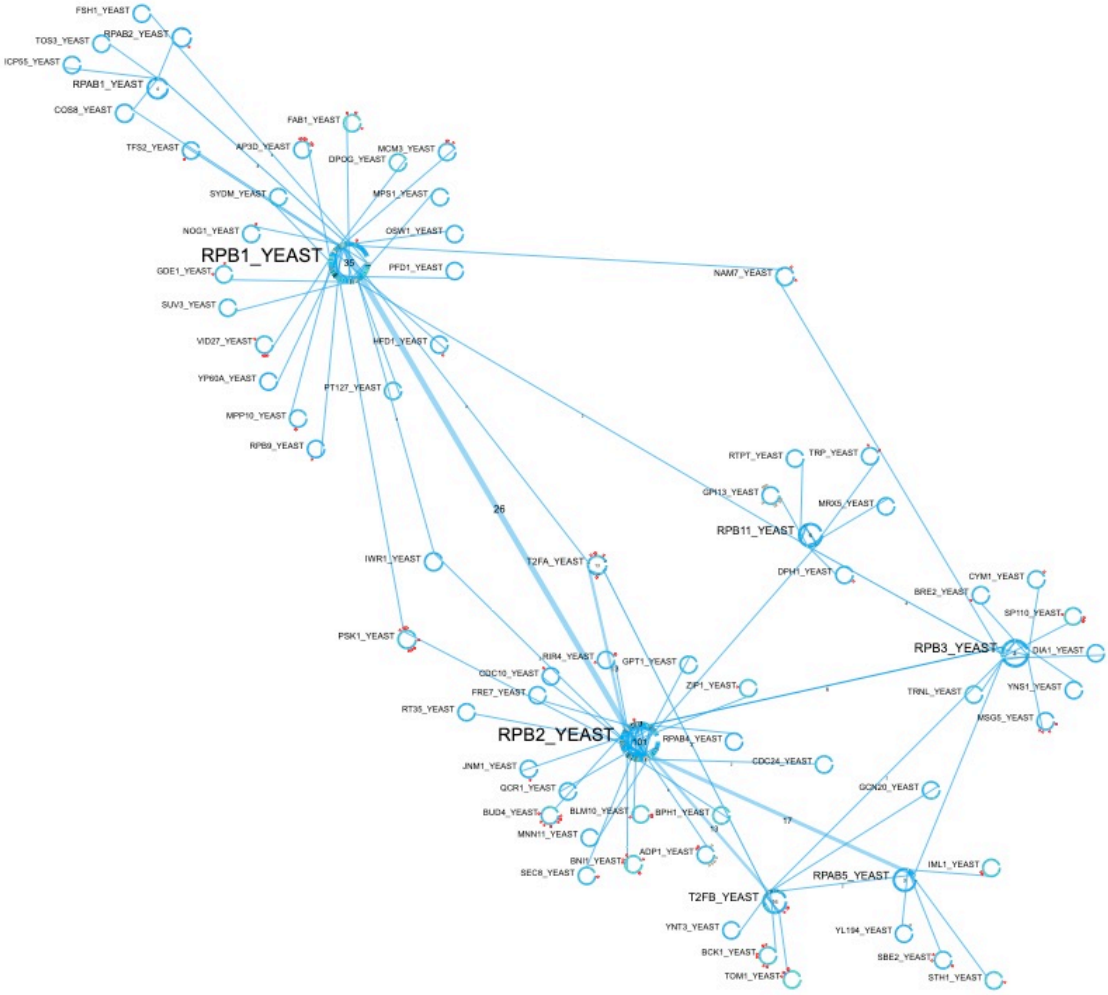


Figure 31: DSSO XL-MS analysis of CF1A and RNAPII complexes by MS2-MS3 CID on an Orbitrap Fusion Lumos. (A) WT CF1A complex mapping intra-links and inter-links (B) WT Rpb3-FLAG mapping intra-links and inter-links.

DISCUSSION

This work explores the role of RNAPII phosphatases Rtr1 and Ssu72 in the regulation of transcription termination and elongation. My work has determined that both phosphatases have distinct roles in regulating transcription termination pathways, specifically NNS-dependent termination. In order to effectively interrogate the complicated mechanism of transcriptional regulation by dynamic RNAPII phosphorylation patterns, several cutting-edge methods were adopted and optimized, such as RNA-Seq, ChIP-exo, AP-MS and XL-MS. Additionally, I was able to use and help develop new methods and approaches in the Mosley lab such as DisCo - a differential PPI approach, and affinity purification-based crosslinking that was instrumental in my studies. Utilization of this multi-omics approach has been critical to the success of my dissertation work.

Rtr1 Regulates Transcription Termination

In the recent past there was a debate as to whether or not Rtr1 was indeed a CTD serine 5 phosphatase. Since Rtr1 does not have the conventional catalytic site found in other classes of protein phosphatases, some questioned whether Rtr1 was a phosphatase. While work from the Mosley lab and others provided strong evidence for its phosphatase activity towards the CTD and other *in vitro* phosphatase assay substrates, the importance of its role *in vivo* remained ambiguous as a consequence of its potential functional redundancy with Ssu72 and its low activity levels in *in vitro* experiments [203, 205, 206,

302, 359, 395, 396]. My work indicates that Rtr1, through its phosphatase activity, has an important role in the regulation of poly-A independent transcription termination.

DisCo analysis, a novel AP-MS method developed in the Mosley Lab, was applied to obtain a detailed snapshot of perturbations in protein-protein interactions during transcription elongation and termination as a result of deleting *RTR1*. The AP-MS SAINT analysis for each of the above purifications revealed that *RTR1* deletion does have an effect on the Nrd1-TAP and Pcf11-FLAG PPIs (Figure 7C). While Nrd1-TAP *rtr1Δ* cells had a low SAINT probability score for interaction with Rpb1, the Pcf11-TAP *rtr1Δ* cells lost moderate SAINT probability score for Rpb1 and Rpb2 interactions. At first glance, the loss of interactions with Pcf11 and gain of interactions with Nrd1 appear to be opposing effects, but it is difficult to interpret exactly how these interaction changes would impact each different termination mechanism. This is especially true since Nrd1 prefers a serine 5 phosphorylated CTD, and Pcf11 prefers a serine 2 phosphorylated CTD. But, both the NNS and CFIa complexes have other subunits which interact with RNA.

Compared to the individual SAINT analysis on the different AP-MS studies, the DisCo study showed that deleting Rtr1 has a global effect. The correlations that different proteins had with each other in the WT decreased and/or disappeared when Rtr1 was removed (Figure 7A/B). It appears that Rtr1 is required to maintain stable interactions which can survive purification of different termination complexes. It is important to note that the DisCo results are not due to the overall decrease between the different complexes of the entire analysis since each correlation value is calculated using data from two bait proteins and the SAINT values from each biological replicate relative to these proteins

(Nrd1 and Pcf11). The decrease of inter-complex interactions suggests that higher order interactions between these complexes are important for proper transcription. Decreased and/or destabilized interactions between complexes could decrease the efficiency of termination. Overall this could reduce diversity in the number of distinct transcripts that are produced during transcription.

How does Rtr1 link the different complexes together? The answer to this question may be found within my genomic analysis of *RTR1* knockout cells. Studying RNA levels is essential to investigate transcription as a surrogate measure of the average activity of the transcription machinery. For this work, total RNA-seq analysis was performed which includes ncRNA (Figure 10). Deletion of *RTR1* showed an overall decrease in total RNA with a greater impact on ncRNA relative to mRNA. This decrease in ncRNA gives the first piece of evidence that *RTR1* deletion may have more of an effect on the NNS dependent termination which is known to regulate ncRNA [12, 148, 153, 165, 376]. The *YKL151C* Northern blots provide an example of a ncRNA *YKL151 AST* having reduced expression, and possibly as a result, an increase in *YKL151* mRNA expression (Figure 15). The single AP-MS SAINT analysis data for Nrd1 in *RTR1* knockout cells supports these findings in that Nrd1 has increased interaction with Rpb1 relative to WT. These data suggest that NNS activity is increased, thus there was a decrease in the number of stable ncRNA detected. This leads to the following hypothesis: deletion of *RTR1* increases S5P levels in the cell which would increase the binding sites for Nrd1 to bind to RNAPII, specifically Rpb1, and increase the efficiency of NNS transcription termination (Figure 32).

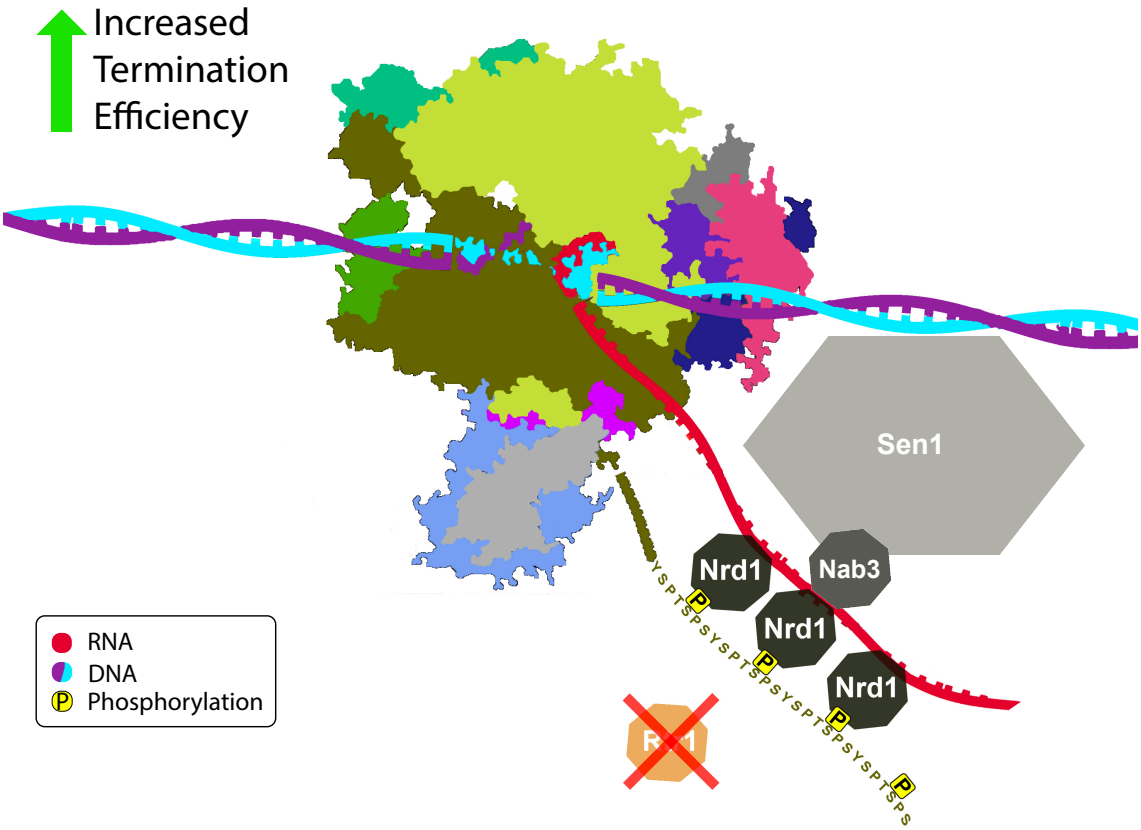


Figure 32: Model for Rtr1 effect on NNS termination. No Rtr1 phosphatase activity leads to increased serine 5 phosphorylated CTD sites. More serine 5 phosphorylated sites lead to increased rate of Nrd1 binding to the CTD. Increased binding of Nrd1 to CTD leads to a more efficient NNS induced RNAPII termination.

Nrd1 and RNAPII occupancy was measured on RNAPII target genes using ChIP-exo (Figure 9-11). The ChIP-exo data show that RNAPII and Nrd1 levels are decreased on the gene as a result of the *RTR1* deletion. It is possible that NNS termination is becoming more efficient from a mechanistic perspective and not as a consequence of a greater number of NNS subunits present with RNAPII, which would mean that less Nrd1 could be required at the termination site to induce termination. When analyzing the changes observed in *rtr1Δ* cells, one should note that the impact of NNS-dependent termination on the transcriptome is regulated by at least two major mechanisms. The first is that the *NRD1* expression levels are autoregulated by the Nrd1 protein itself. Thus, an increase in Nrd1 termination efficiency would lead to a corresponding decrease in *NRD1* mRNA and possibly protein levels. Secondly, increased Nrd1 termination could be causing a reduced number of stable ncRNA in the cell which would be mediated through the activity of Rrp6 in the nuclear RNA exosome. It is possible that RNAPII could be transcribing the ncRNA properly, but if Rrp6 is acting more efficiently, it could be degrading the RNA before it is measured with my RNA-Seq experiments. For this reason, an *RRP6* knockout mutant and an *RTR1 RRP6* double knockout mutant were used to assess whether degradation of the ncRNA alone could explain the results. Two examples of well-studied protein coding genes that are known to be regulated by transcription of ncRNA through transcription interference are shown. Analysis of both *IMD2* and *YKL151C* do not support the hypothesis that the effect caused by deleting *RTR1* leads the RNA exosome to be defective and degrade ncRNA. As assessed by RNA-Seq and Northern blot analysis, the *rrp6Δ* cells differ from the *rtr1Δ/rrp6Δ*. If the effect of *RTR1* deletion was dependent on Rrp6, then the double mutant should act identically to cells with only *rrp6Δ*. However,

when looking at global AST transcripts, the double mutant did resemble the single *RRP6* knockout mutant. This would mean that most AST transcripts do depend on Rrp6 activity and that Rtr1 does have an influence on the RNA exosome on most genes. Collectively these results lead me to conclude that Rtr1 may play a role in attenuating NNS dependent termination. However, it also plays a role in positively regulating the stability of ncRNA via the RNA exosome. This is consistent with previous work in the Mosley lab that showed that Rrp6 acts with the NNS complex to promote transcription termination as well as processing and degradation[248].

Ssu72-L84F Mutant Elucidates Possible Roles in Transcription Termination Mechanism That Are Independent From Rtr1s Role

Ssu72 was the first S5P phosphatase discovered, and, because of its interactions with TFIIB, has commanded significant attention in the field. As an essential member of the CPF, it is likely involved in poly-A dependent termination and possibly also in initiation through its interaction with TFIIB. In addition, as a S5P phosphatase, it should also theoretically be involved in early elongation when S5P is most abundant [180, 393]. The role of Rtr1 and Ssu72 phosphatases in RNAPII transcription and how they differ and are similar have not been fully defined. Investigation of these areas is important to determine whether Rtr1 and Ssu72 are redundant and/or specific. There is evidence that Ssu72 prefers a serine 5 phosphorylated CTD where, a proline 6 is in a trans conformation [303]. However, other conformational differences and combinations of phosphorylated residues, in addition to proline 6 isomerization, could exist that each of the two phosphatases prefer.

Ssu72 is an essential gene, as thus most studies have focused on an Ssu72 mutant. All of these mutants have generally caused a loss of function in Ssu72 that led to defects in transcription. One such study found a set of transcripts, that were affected by a Ssu72-R129A mutant, which they called the SRTs [13]. In this study I used a new mutant, Ssu72-L84F, which arose spontaneously in a genetic screen that appeared to over-ride termination [350]). As the other Ssu72 mutants were well studied and could already be compared to my Rtr1 data, I wanted to look at the new Ssu72-L84F mutant to have an additional unique set of Ssu72 mutant data to compare to Rtr1 as well as to better understand what effect the mutation of Ssu72 had on Ssu72s activity and in the cell.

In three separate experiments, it was confirmed that Ssu72-L84F has increased phosphatase activity, which is the first Ssu72 mutant to have such a phenotype (Figure 16). Interestingly it appears as though other mutations in the same location would have even greater effects, but it is uncertain whether a yeast cell with this mutation would be viable or even have the same TOV phenotype. In addition, all of the phosphorylation levels at the serine residues in the CTD appeared to be decreased although it is not clear if they are directly dephosphorylated by Ssu72 or by a mechanism dependent on Ssu72. Since Ssu72 is known to also dephosphorylate S7P and is also a member of the phosphatase module in the CPF, S2/5/7P dephosphorylation could be a combination of both direct dephosphorylation of the serine residues and stimulation of their dephosphorylation through downstream mechanisms, such as Fcp1 activity.

Total RNA-seq analysis of Ssu72-L84F cells was performed to compare the differential expression analysis of the two phosphatases directly without experimental bias. Again, in the Ssu72-L84F data it was the ncRNA that were affected, but this time

the ncRNA were all increased very distinctly relative to *rtr1Δ* cells that had a decrease in the ncRNA expression levels. Curiously, it appears that different classes of ncRNA were affected more by Rtr1 (ASTs) and others were affected more by Ssu72 mutations (CUTS, SRTs, and SNRs) while both seemed to affect NUTs in a similar way. Northern blot analysis was also performed to confirm genuine read-through and the presence of their full-length transcripts of SNR genes snR33 and snr82. Curiously the levels of RNA expression of the elongated transcripts relative to the correct snoRNA as measured by RNA-seq and Northern blot were not in agreement. Northern blot data displayed a much smaller elongated transcript blot relative to the short snoRNA transcripts. Whereas the RNA-seq data displayed similar levels of elongated vs snoRNA expression levels. It is possible that this difference is result of the limitations of RNA-seq when measuring very high quantities of RNA as they could have a saturation maximum or problems with library amplification.

Analysis of RNA expression in an *rrp6Δ* mutant showed that the Ssu72-L84F mutation effect was not an effect of a defective RNA exosome since the *rrp6Δ* lane showed a different shorter RNA transcript product. The difference in RNA products show further evidence that the Ssu72-L84F mutation has a termination read-through effect and not a degradation defect phenotype. The read-through transcript size aligns with the length of DNA from the SNRs TSS to the downstream genes TES which means that the Ssu72-L84F mutation is affecting ncRNA termination but probably not poly-A dependent termination. It is interesting to note that when a double *rtr1Δ* Ssu72-L84F mutant was used, Northern blot analysis displayed the characteristic of Ssu72-L84F in that it there was a very similar read-through effect (data not shown). This is evidence that the L84F

mutation on Ssu72 could be inducing a loss of its specificity and targeting S5P that would have normally been dephosphorylated. This hypothesis makes sense since the L84F mutation is localized near the de-phosphorylation active site and loss of specificity would allow Ssu72 to dephosphorylate indiscriminately at a higher rate.

ChIP-exo results of an Rpb3-FLAG tagged strain for WT and Ssu72-L84F for snR82 and snR33 displayed similar results as the RNA-seq analysis when comparing the mutant to the WT, although the difference between the two strains in the RNA-seq analysis was much more pronounced. The global average reads from ChIP-exo data for RNAPII showed more interesting results in both ncRNA and mRNA genes. The differences between the average RNAPII occupancy for Ssu72-L84F compared to WT at snRNA genes was quite pronounced at both the TES and the TSS. At the TSS there is not much difference in RNAPII occupancy which gives evidence to the idea that the mutant does not affect initiation of transcription. It also appears as though elongation is not affected in the SNR genes. Whereas the SNR genes have a large shift downstream of the TES relative to WT which is consistent with the idea that the Ssu72-L84F mutant is causing termination over-ride at the SNR genes. In addition, there was an accumulation of RNAPII at the TES in the SNR genes, and it appears as though RNAPII is stalling at this region, perhaps trying to initiate termination but for some reason being unable. This same accumulation of RNAPII at the TES is also apparent in the protein coding genes. The shift downstream of the TES is not present which could mean that the Ssu72-L84F mutation is causing an accumulation of RNAPII at the TES of poly-A dependent genes but is still terminating at the appropriate location indicating that it might be temporally delayed. Surprisingly, there was also a decrease in RNAPII occupancy during early

elongation of protein coding genes, an effect that was not seen in the SNR genes. My model postulates that the Ssu72-L84F mutation is having an effect on NNS termination which occurs in a similar area as peak Nrd1-TAP occupancy, which can be seen when compared to the average Rpb3-FLAG occupancy ChIP-exo data (Figure 33). As mentioned, the localization of Nrd1 is at the same region where RNAPII localization is reduced in the protein coding genes for the Ssu72-L84F mutant. This result leads to the model that Ssu72-L84F mutation is increasing phosphatase activity thus reducing the sites of S5P that Nrd1 can bind to and inhibit NNS dependent termination (Figure 34). Therefore, I believe that the Ssu72-L84F mutation is affecting NNS termination of ncRNA, but the NNS complex might also be involved in early elongation in protein coding genes as a possible checkpoint that RNAPII has to go through before it continues to elongate the nascent RNA.

The Ssu72-L84F mutant had a very drastic effect on RNAPII localization, but the effects of *rtr1Δ* on RNAPII were generally subtler. It appears as though Ssu72 phosphatase is necessary to stimulate NNS termination, but Rtr1 actually attenuates NNS termination. Although Ssu72 has been shown to be involved in termination previously, this is a novel role for Rtr1. It is interesting that both Ssu72 and Rtr1 seemingly target the same serine residue on the CTD, but they each appear to have opposite effects on transcription. One could say that the mutation confers opposite phenotypes on the phosphatases themselves. However, other mutations in Ssu72 that have decreased phosphatase activity also confer a read-through phenotype [13]. This gives further evidence that these phosphatases have a separate and unique role in transcription and that they are probably targeting structurally unique RNAPII S5P sites.

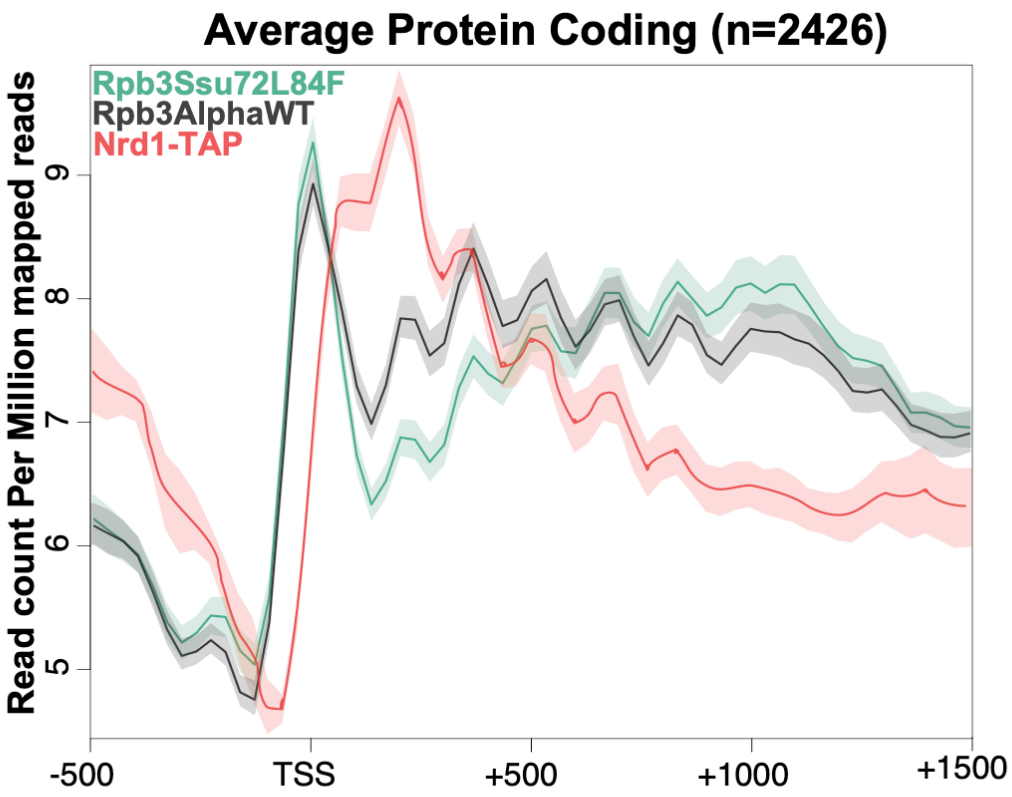


Figure 33: Global ChIP-exo analysis using NGS.plot program. Non-overlapping protein coding genes were analyzed. Graph is centered around transcription start sites of each gene. Read counts are averaged across genes in the annotation list and plotted. TSS of protein coding genes show a more interesting picture. Nrd1-TAP ChIP-exo data is overlaid to the Ssu72-L84F data. The drop in RNAPII localization during early elongation aligns with spike of localization of Nrd1. This further supports my evidence that Ssu72 and the NNS pathway might be affecting protein coding genes.

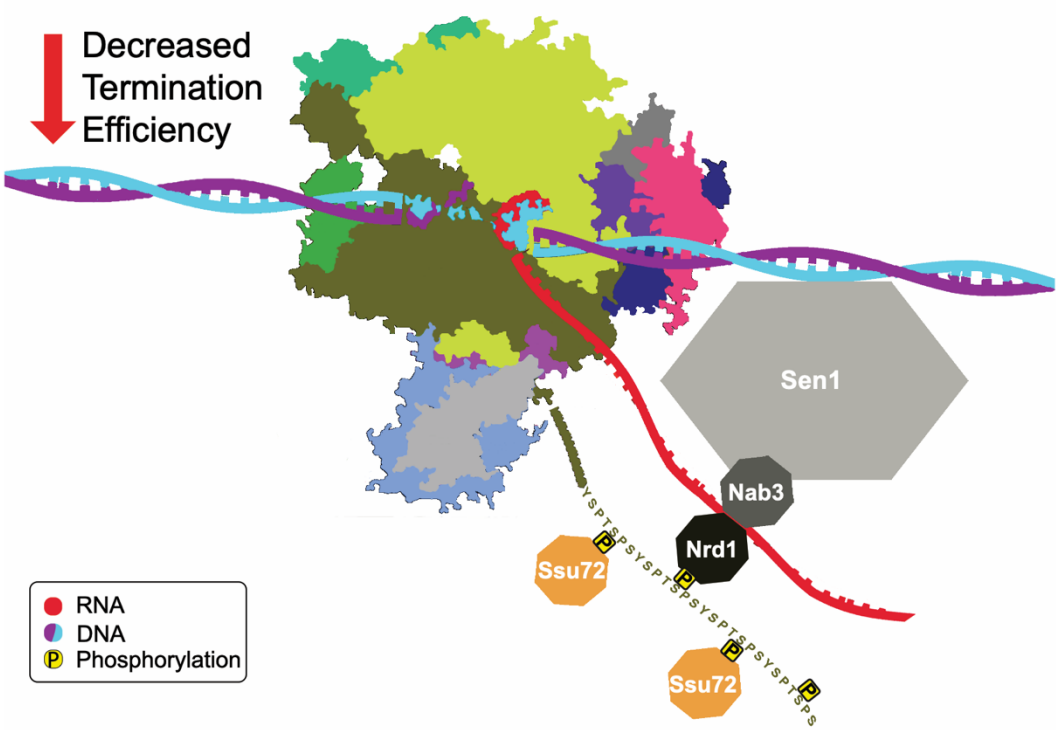


Figure 34: Model for Ssu72-L84F effect on NNS termination. Increased phosphatase activity of Ssu72 leads to fewer serine 5 phosphorylated CTD sites. Less serine 5 phosphorylated sites lead to decreased rate of Nrd1 binding to CTD. Decreased binding of NRd1 to CTD leads to a less efficient NNS induced RNAPII termination and less Nrd1 induced RNAPII pausing.

This hypothesis could be further corroborated by studies with a hyperactive Rtr1 if such a scenario could be engineered.

Ssu72-L84F Reveals Independent Functionality in Subcomplex Within CPF

Our Ssu72-L84F studies revealed possible independent functionalities of a subcomplex within the CPF complex. It is possible that Ssu72s main function in transcription is further downstream of the TSS relative to Rtr1. The mutation may have caused a decrease in Ssu72 dephosphorylation site specificity causing the termination over-ride phenotype. However, differential PPI proteomics data showed a distinct separation of a group of different subunits within the CPF complex due to the Ssu72-L84F mutation. Furthermore, this same group of proteins has recently been described as an independent group structurally separate from the rest of the CPF in a different experiment. These data may suggest the functionality of this group using the Ssu72-L84F mutant. It is possible that this subcomplex of proteins is acting separately from the rest of the complex, which is supported by recent work showing that when the Syc1 protein is present there exists a distinct subcomplex APT. APT includes Ssu72 as well as the proteins that have differential PPI in the Ssu72-L84F strains. However, my differential PPI data does not include all of the APT complex members, on at least not with a high SAINT probability. It is possible that that a smaller subcomplex exists within the APT complex that has a distinct function within the cell that is being affected by the Ssu72-L84F mutation. For this reason, the architecture of the protein complexes in the Ssu72-L84F mutation were investigated through the development of the cross-linking mass spectrometry methods. The cross-linking method shows promise and with further

development this method should become a powerful tool for analyzing the effects that the Ssu72-L84F mutation as well as other mutations have on their respective complexes.

Conclusions and Future Directions

Through this work I have been able to distinguish the unique roles that the Rtr1 and Ssu72 phosphatase have on transcription termination. However, this work has also brought about new questions that would be interesting to answer. My overall findings have once again displayed that biology is extremely complicated system and that even the most seemingly simple mechanisms are riddled with intricate details is connected to several aspects of biology.

Rtr1 was recently (2009) a protein of unknown function. Even more recently its function was still debated. However, work in the Mosley lab and others have uncovered more information about the protein and its function. In this work my data provides evidence the Rtr1 appears to be acting as a sort of gateway that enables elongation to continue and promotes protein interactions between transcription machinery. Most Ssu72 mutants, loss of phosphatase activity mutants or gain of phosphatase activity mutants, have reduced termination and increased readthrough which is the opposite of the *rtr1Δ* strain even though both are S5P phosphatases. This result then introduces the question that if removing Rtr1 has the same effect on serine 5 of the CTD as a loss of activity Ssu72 mutant phosphatase then why is the phenotype different. Again, the simple action of removing the S5P from the CTD is becoming increasingly intricate and my results provide evidence that this is likely the case for the other phosphatases and kinases as well as the other transcription machinery.

It would be interesting to verify whether or not the *rtr1Δ* phenotype is a direct result of the loss of S5P. If it is then it would be worthwhile to investigate the structural differences between an Rtr1 S5P CTD target and an Ssu72 S5P target. This could possibly be done by using structurally unique synthetic heptad peptides of various repeats, phosphorylation patterns, and other structurally unique peptides in a phosphatase activity assay. If the *rtr1Δ* phenotype is a result of something other than the loss of S5P then it is possible that Rtr1 may be acting on other proteins involved in transcription and further studies can be done to look at differential phospho-proteomics as a result of the *rtr1Δ* to identify possible targets of Rtr1. In addition, I have preliminary Northern blot data that shows that an *rtr1Δ/*Ssu72-L84F double mutant results in RNA products similar to those found in the Ssu72-L84F mutant so it appears that the Ssu72 mutations overwhelms the effects of *RTR1* deletion. However, it is still unclear if the double mutant reduces the interactions between the transcription machinery in the same way as an *rtr1Δ* strain. Thus, in future studies it would be a good idea to implement a DisCo study in the double mutant to see if the interactions between the different transcription machinery are restored.

My studies with Ssu72 led to surprising results in its effect in early mRNA elongation and within the complex that it is a part of CPF. Not only did the Ssu72-L84F mutation confer a definitive read-through effect in ncRNA genes but they also led to a decrease in occupancy at the early elongation stage and an accumulation in effects in mRNA genes. The differential PPI analysis identified a subset of the APT module proteins are interacting with Pta1 more often than other CPF members. The finding that Ssu72 is not only acting during at termination with its complex but within a sub-complex

of the module again highlights the number of layers that exist in transcription mechanism that is occurring within the cell. It also calls to attention the heterogeneity that protein complexes have. Proteins may not always be in just one complex but they may act independently or in a completely separate complex.

In the future I will continue the cross-linking studies to gain further information on the interactions that occur in CPF, CF1A, as well as other important protein complexes involved in transcription. By doing this I can create a very detailed map of how proteins interact with each other going one step further than AP-MS PPI studies that cannot distinguish between primary, secondary or n interactions. In addition, I will be able to distinguish if there is a change in which regions of the protein where they are interacting as a result of the Ssu72-L84F or other mutations. I will also perform a whole proteome experiment using tandem mass tags to accurately quantify the amount of different proteins in the cell to find out if there is a change in the expression of different proteins. In order to verify that there is a sub complex within the APT module of the CPF in the Ssu72-L84F mutant I will do an anion exchange separation experiment with the CPF complex to identify if a different sub-complex exists on its own and travels together in the column. Since RNAPII is accumulated at the 3' end of mRNA genes in the Ssu72-L84F mutant relative to WT it is possible that cleavage and polyadenylation function is altered. To test this, I will perform a cleavage and polyadenylation assay using an RNA substrate with purified CPF and CF1A complex from an Ssu72-L84F mutant strain as well as the WT.

REFERENCES

1. Lee, T.I. and R.A. Young, *Transcriptional regulation and its misregulation in disease.* Cell, 2013. **152**(6): p. 1237-51.
2. Guglielmi, V., et al., *Abnormal expression of RNA polymerase II-associated proteins in muscle of patients with myofibrillar myopathies.* Histopathology, 2015. **67**(6): p. 859-865.
3. Kim, S.H., et al., *Hepatocyte homeostasis for chromosome ploidization and liver function is regulated by Ssu72 protein phosphatase.* Hepatology, 2016. **63**(1): p. 247-59.
4. Chen, Y., et al., *A gene-specific role for the Ssu72 RNAPII CTD phosphatase in HIV-1 Tat transactivation.* Genes Dev, 2014. **28**(20): p. 2261-75.
5. Chen, Y.Z., et al., *Senataxin, the yeast Sen1p orthologue: Characterization of a unique protein in which recessive mutations cause ataxia and dominant mutations cause motor neuron disease.* Neurobiology of Disease, 2006. **23**(1): p. 97-108.
6. Richard, P., S. Feng, and J.L. Manley, *A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage.* Genes & Development, 2013. **27**(20): p. 2227-2232.
7. Boczonadi, V., et al., *EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia.* Nature Communications, 2014. **5**.
8. Bradner, J.E., D. Hnisz, and R.A. Young, *Transcriptional Addiction in Cancer.* Cell, 2017. **168**(4): p. 629-643.
9. Cramer, P., *Multisubunit RNA polymerases.* Current Opinion in Structural Biology, 2002. **12**(1): p. 89-97.
10. Vannini, A. and P. Cramer, *Conservation between the RNA polymerase I, II, and III transcription initiation machineries.* Mol Cell, 2012. **45**(4): p. 439-46.
11. Warner, J.R., *The economics of ribosome biosynthesis in yeast.* Trends in Biochemical Sciences, 1999. **24**(11): p. 437-440.
12. Wyers, F., et al., *Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase.* Cell, 2005. **121**(5): p. 725-37.
13. Tan-Wong, S.M., et al., *Gene loops enhance transcriptional directionality.* Science, 2012. **338**(6107): p. 671-5.
14. Sydow, J.F., et al., *Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA.* Mol Cell, 2009. **34**(6): p. 710-21.
15. Edwards, A.M., et al., *Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro.* J Biol Chem, 1991. **266**(1): p. 71-5.
16. Mosley, A.L., et al., *Quantitative proteomics demonstrates that the RNA polymerase II subunits Rpb4 and Rpb7 dissociate during transcriptional elongation.* Mol Cell Proteomics, 2013. **12**(6): p. 1530-8.

17. Cramer, P., et al., *Architecture of RNA polymerase II and implications for the transcription mechanism.* Science, 2000. **288**(5466): p. 640-9.
18. Sainsbury, S., C. Bernecky, and P. Cramer, *Structural basis of transcription initiation by RNA polymerase II.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 129-43.
19. Basehoar, A.D., S.J. Zanton, and B.F. Pugh, *Identification and distinct regulation of yeast TATA box-containing genes (vol 116, pg 699, 2004).* Cell, 2004. **117**(6): p. 847-847.
20. Rhee, H.S. and B.F. Pugh, *Genome-wide structure and organization of eukaryotic pre-initiation complexes.* Nature, 2012. **483**(7389): p. 295-301.
21. Baumann, M., J. Pontiller, and W. Ernst, *Structure and Basal Transcription Complex of RNA Polymerase II Core Promoters in the Mammalian Genome: An Overview.* Molecular Biotechnology, 2010. **45**(3): p. 241-247.
22. Yang, C.H., et al., *Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters.* Gene, 2007. **389**(1): p. 52-65.
23. Imbalzano, A.N., K.S. Zaret, and R.E. Kingston, *Transcription Factor (Tf) Iib and Tfiia Can Independently Increase the Affinity of the Tata-Binding Protein for Pna.* Journal of Biological Chemistry, 1994. **269**(11): p. 8280-8286.
24. Zhao, X.M. and W. Herr, *A regulated two-step mechanism of TBP binding to DNA: A solvent-exposed surface of TBP inhibits TATA box recognition.* Cell, 2002. **108**(5): p. 615-627.
25. Ha, I., W.S. Lane, and D. Reinberg, *Cloning of a Human Gene Encoding the General Transcription Initiation Factor-Iib.* Nature, 1991. **352**(6337): p. 689-695.
26. Buratowski, S. and H. Zhou, *Functional Domains of Transcription Factor-Tfiib.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(12): p. 5633-5637.
27. Conaway, R.C., et al., *Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex.* Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6205-9.
28. Conaway, R.C. and J.W. Conaway, *An Rna Polymerase-Ii Transcription Factor Has an Associated DNA-Dependent Atpase (Datpase) Activity Strongly Stimulated by the Tata Region of Promoters.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(19): p. 7356-7360.
29. Flores, O., E. Maldonado, and D. Reinberg, *Factors Involved in Specific Transcription by Mammalian Rna Polymerase-Ii - Factor-Iie and Factor-Iif Independently Interact with Rna Polymerase-Ii.* Journal of Biological Chemistry, 1989. **264**(15): p. 8913-8921.
30. Maxon, M.E., J.A. Goodrich, and R. Tjian, *Transcription Factor Iie Binds Preferentially to Rna-Polymerase Iia and Recruits Tfiih - a Model for Promoter Clearance.* Genes & Development, 1994. **8**(5): p. 515-524.
31. Schaeffer, L., et al., *The Ercc2/DNA Repair Protein Is Associated with the Class-Ii Btf2/Tfiih Transcription Factor.* Embo Journal, 1994. **13**(10): p. 2388-2392.
32. Feaver, W.J., et al., *Ctd Kinase Associated with Yeast Rna Polymerase-Ii Initiation Factor-B.* Cell, 1991. **67**(6): p. 1223-1230.

33. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7.* Science, 2007. **318**(5857): p. 1780-1782.
34. Kim, M., et al., *Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7.* Journal of Biological Chemistry, 2009. **284**(39): p. 26421-26426.
35. Akhtar, M.S., et al., *TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II.* Mol Cell, 2009. **34**(3): p. 387-93.
36. Glover-Cutter, K., et al., *TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II.* Mol Cell Biol, 2009. **29**(20): p. 5455-64.
37. Schaeffer, L., et al., *DNA-Repair Helicase - a Component of Btf2 (Tfiih) Basic Transcription Factor.* Science, 1993. **260**(5104): p. 58-63.
38. Holstege, F.C.P., P.C. vanderVliet, and H.T.M. Timmers, *Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH.* Embo Journal, 1996. **15**(7): p. 1666-1677.
39. Goodrich, J.A. and R. Tjian, *Transcription Factors Iie and Iih and Atp Hydrolysis Direct Promoter Clearance by Rna-Polymerase-Ii.* Cell, 1994. **77**(1): p. 145-156.
40. Moreland, R.J., et al., *A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II.* J Biol Chem, 1999. **274**(32): p. 22127-30.
41. Egly, J.M. and F. Coin, *A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor.* DNA Repair (Amst), 2011. **10**(7): p. 714-21.
42. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 angstrom resolution.* Nature, 1997. **389**(6648): p. 251-260.
43. Ehara, H., et al., *Structural insight into nucleosome transcription by RNA polymerase II with elongation factors.* Science, 2019. **363**(6428): p. 744-+.
44. Gates, L.A., C.E. Foulds, and B.W. O'Malley, *Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle.* Trends in Biochemical Sciences, 2017. **42**(12): p. 977-989.
45. Rhee, H.S., et al., *Subnucleosomal structures and nucleosome asymmetry across a genome.* Cell, 2014. **159**(6): p. 1377-88.
46. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.
47. Rando, O.J. and K. Ahmad, *Rules and regulation in the primary structure of chromatin.* Current Opinion in Cell Biology, 2007. **19**(3): p. 250-256.
48. Zhang, L.Y., H. Ma, and B.F. Pugh, *Stable and dynamic nucleosome states during a meiotic developmental process.* Genome Research, 2011. **21**(6): p. 875-884.
49. Liu, Y., et al., *Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes.* Journal of Biological Chemistry, 2001. **276**(10): p. 7169-7175.
50. Esnault, C., et al., *Mediator-dependent recruitment of TFIIH modules in preinitiation complex.* Molecular Cell, 2008. **31**(3): p. 337-346.
51. Corden, J.L., *RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template.* Chemical Reviews, 2013. **113**(11): p. 8423-8455.
52. Allen, B.L. and D.J. Taatjes, *The Mediator complex: a central integrator of transcription.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 155-66.

53. Naar, A.M., et al., *Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation.* Genes Dev, 2002. **16**(11): p. 1339-44.
54. Donner, A.J., et al., *CDK8 is a positive regulator of transcriptional elongation within the serum response network.* Nature Structural & Molecular Biology, 2010. **17**(2): p. 194-U9.
55. Ebmeier, C.C. and D.J. Taatjes, *Activator-Mediator binding regulates Mediator-cofactor interactions.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(25): p. 11283-11288.
56. Lu, H.S., et al., *AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(1): p. E15-E24.
57. Zhou, Q., T. Li, and D.H. Price, *RNA polymerase II elongation control.* Annu Rev Biochem, 2012. **81**: p. 119-43.
58. Lin, C.Q., et al., *Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC).* Genes & Development, 2011. **25**(14): p. 1486-1498.
59. Luo, Z., C. Lin, and A. Shilatifard, *The super elongation complex (SEC) family in transcriptional control.* Nat Rev Mol Cell Biol, 2012. **13**(9): p. 543-7.
60. Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* Science, 2008. **322**(5909): p. 1845-8.
61. Yamaguchi, Y., et al., *NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation.* Cell, 1999. **97**(1): p. 41-51.
62. Vos, S.M., et al., *Architecture and RNA binding of the human negative elongation factor.* Elife, 2016. **5**.
63. Schaukowitch, K., et al., *Enhancer RNA facilitates NELF release from immediate early genes.* Mol Cell, 2014. **56**(1): p. 29-42.
64. Jadlowsky, J.K., et al., *Negative Elongation Factor Is Required for the Maintenance of Proviral Latency but Does Not Induce Promoter-Proximal Pausing of RNA Polymerase II on the HIV Long Terminal Repeat.* Molecular and Cellular Biology, 2014. **34**(11): p. 1911-1928.
65. Core, L. and K. Adelman, *Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation.* Genes Dev, 2019. **33**(15-16): p. 960-982.
66. Hirtreiter, A., et al., *Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif.* Nucleic Acids Research, 2010. **38**(12): p. 4040-4051.
67. Kyrpides, N.C., C.R. Woese, and C.A. Ouzounis, *KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins.* Trends Biochem Sci, 1996. **21**(11): p. 425-6.
68. Guo, M., et al., *Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation.* Structure, 2008. **16**(11): p. 1649-58.

69. Qiu, Y. and D.S. Gilmour, *Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing.* J Biol Chem, 2017. **292**(13): p. 5555-5570.
70. Booth, G.T., et al., *Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast.* Nat Commun, 2018. **9**(1): p. 543.
71. Parua, P.K., et al., *A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II.* Nature, 2018. **558**(7710): p. 460-464.
72. Wen, Y. and A.J. Shatkin, *Transcription elongation factor hSPT5 stimulates mRNA capping.* Genes Dev, 1999. **13**(14): p. 1774-9.
73. Perales, R. and D. Bentley, *"Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions.* Molecular Cell, 2009. **36**(2): p. 178-191.
74. Izaurralde, E., et al., *A Cap-Binding Protein Complex Mediating U Snrna Export.* Nature, 1995. **376**(6542): p. 709-712.
75. Rasmussen, E.B. and J.T. Lis, *In-Vivo Transcriptional Pausing and Cap Formation on 3 Drosophila Heat-Shock Genes.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(17): p. 7923-7927.
76. Peck, S.A., et al., *Writing a wrong: Coupled RNA polymerase II transcription and RNA quality control.* Wiley Interdiscip Rev RNA, 2019. **10**(4): p. e1529.
77. Tsukamoto, T., et al., *Isolation and characterization of the yeast mRNA capping enzyme beta subunit gene encoding RNA 5'-triphosphatase, which is essential for cell viability.* Biochemical and Biophysical Research Communications, 1997. **239**(1): p. 116-122.
78. Rodriguez, C.R., et al., *A Saccharomyces cerevisiae RNA 5 '-triphosphatase related to mRNA capping enzyme.* Nucleic Acids Research, 1999. **27**(10): p. 2181-2188.
79. Shibagaki, Y., et al., *Messenger-Rna Capping Enzyme - Isolation and Characterization of the Gene Encoding Messenger-Rna Guanylyltransferase Subunit from Saccharomyces-Cerevisiae.* Journal of Biological Chemistry, 1992. **267**(14): p. 9521-9528.
80. Mao, X.D., B. Schwer, and S. Shuman, *Yeast Messenger-Rna Cap Methyltransferase Is a 50-Kilodalton Protein Encoded by an Essential Gene.* Molecular and Cellular Biology, 1995. **15**(8): p. 4167-4174.
81. Glover-Cutter, K., et al., *RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes.* Nat Struct Mol Biol, 2008. **15**(1): p. 71-8.
82. Martinez-Rucobo, F.W., et al., *Molecular Basis of Transcription-Coupled Pre-mRNA Capping.* Mol Cell, 2015. **58**(6): p. 1079-89.
83. McCracken, S., et al., *5 '-capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II.* Genes & Development, 1997. **11**(24): p. 3306-3318.
84. Izaurralde, E., et al., *A Nuclear Cap-Binding Protein Complex Involved in Pre-Messenger-Rna Splicing.* Cell, 1994. **78**(4): p. 657-668.

85. Izaurralde, E., et al., *A Cap Binding-Protein That May Mediate Nuclear Export of Rna-Polymerase Ii-Transcribed Rnas.* Journal of Cell Biology, 1992. **118**(6): p. 1287-1295.
86. Schmid, M. and T.H. Jensen, *Controlling nuclear RNA levels.* Nat Rev Genet, 2018. **19**(8): p. 518-529.
87. Jimeno-Gonzalez, S., et al., *The Yeast 5 '-3 ' Exonuclease Rat1p Functions during Transcription Elongation by RNA Polymerase II.* Molecular Cell, 2010. **37**(4): p. 580-587.
88. Oesterreich, F.C., et al., *Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II.* Cell, 2016. **165**(2): p. 372-381.
89. Pandya-Jones, A. and D.L. Black, *Co-transcriptional splicing of constitutive and alternative exons.* Rna, 2009. **15**(10): p. 1896-1908.
90. Kim, J., M. Guermah, and R.G. Roeder, *The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS.* Cell, 2010. **140**(4): p. 491-503.
91. Chen, F.X., et al., *PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II.* Cell, 2015. **162**(5): p. 1003-15.
92. Van Oss, S.B., et al., *The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6.* Molecular Cell, 2016. **64**(4): p. 815-825.
93. Day, D.S., et al., *Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types.* Genome Biology, 2016. **17**.
94. Kim, J.B. and P.A. Sharp, *Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase.* J Biol Chem, 2001. **276**(15): p. 12317-23.
95. Marshall, N.F., et al., *Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase.* Journal of Biological Chemistry, 1996. **271**(43): p. 27176-27183.
96. Fujinaga, K., et al., *Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element.* Mol Cell Biol, 2004. **24**(2): p. 787-95.
97. Hondele, M., et al., *Structural basis of histone H2A-H2B recognition by the essential chaperone FACT.* Nature, 2013. **499**(7456): p. 111-4.
98. Liu, Y., et al., *FACT caught in the act of manipulating the nucleosome.* Nature, 2019.
99. Petesch, S.J. and J.T. Lis, *Overcoming the nucleosome barrier during transcript elongation.* Trends Genet, 2012. **28**(6): p. 285-94.
100. Krogan, N.J., et al., *The Paf1 complex is required for histone h3 methylation by COMPASS and Dot1p: Linking transcriptional elongation to histone methylation.* Molecular Cell, 2003. **11**(3): p. 721-729.
101. Wood, A., et al., *The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p.* Journal of Biological Chemistry, 2003. **278**(37): p. 34739-34742.

102. Simic, R., et al., *Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes.* Embo Journal, 2003. **22**(8): p. 1846-1856.
103. Farnung, L., S.M. Vos, and P. Cramer, *Structure of transcribing RNA polymerase II-nucleosome complex.* Nat Commun, 2018. **9**(1): p. 5432.
104. Vos, S.M., et al., *Structure of activated transcription complex Pol II-DSIF-PAF-SPT6.* Nature, 2018. **560**(7720): p. 607-612.
105. Kessler, M.M., et al., *Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast.* Genes & Development, 1997. **11**(19): p. 2545-2556.
106. Chen, J. and C. Moore, *Separation of Factors Required for Cleavage and Polyadenylation of Yeast Pre-Messenger-Rna.* Molecular and Cellular Biology, 1992. **12**(8): p. 3470-3481.
107. Birse, C.E., et al., *Coupling termination of transcription to messenger RNA maturation in yeast.* Science, 1998. **280**(5361): p. 298-301.
108. Fatica, A., M. Morlando, and I. Bozzoni, *Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus.* EMBO J, 2000. **19**(22): p. 6218-29.
109. Licatalosi, D.D., et al., *Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II.* Mol Cell, 2002. **9**(5): p. 1101-11.
110. Ryan, K., O. Calvo, and J.L. Manley, *Evidence that polyadenylation factor CPSF-73 is the rnRNA 3 ' processing endonuclease.* Rna, 2004. **10**(4): p. 565-573.
111. Christofori, G. and W. Keller, *3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP.* Cell, 1988. **54**(6): p. 875-89.
112. Zhao, J., et al., *Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor.* Mol Cell Biol, 1999. **19**(11): p. 7733-40.
113. Mandel, C.R., Y. Bai, and L. Tong, *Protein factors in pre-mRNA 3'-end processing.* Cellular and Molecular Life Sciences, 2008. **65**(7-8): p. 1099-1122.
114. Proweller, A. and S. Butler, *Efficient translation of poly(A)-deficient mRNAs in Saccharomyces cerevisiae.* Genes Dev, 1994. **8**(21): p. 2629-40.
115. Truant, R., et al., *Identification and functional characterization of a novel nuclear localization signal present in the yeast Nab2 Poly(A)(+) RNA binding protein.* Molecular and Cellular Biology, 1998. **18**(3): p. 1449-1458.
116. West, S., N. Gromak, and N.J. Proudfoot, *Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites.* Nature, 2004. **432**(7016): p. 522-5.
117. Kim, M., et al., *The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II.* Nature, 2004. **432**(7016): p. 517-22.
118. Lunde, B.M., et al., *Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain.* Nature Structural & Molecular Biology, 2010. **17**(10): p. 1195-+.
119. Luo, W., A.W. Johnson, and D.L. Bentley, *The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model.* Genes Dev, 2006. **20**(8): p. 954-65.

120. Dengl, S. and P. Cramer, *Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro.* Journal of Biological Chemistry, 2009. **284**(32): p. 21270-21279.
121. Xiang, S., et al., *Structure and function of the 5 '-> 3 ' exoribonuclease Rat1 and its activating partner Rai1.* Nature, 2009. **458**(7239): p. 784-U130.
122. Jiao, X.F., et al., *Identification of a quality-control mechanism for mRNA 5 '-end capping.* Nature, 2010. **467**(7315): p. 608-U137.
123. Chang, J.H., et al., *Dxo1 is a new type of eukaryotic enzyme with both decapping and 5 '-3 ' exoribonuclease activity.* Nature Structural & Molecular Biology, 2012. **19**(10): p. 1011-U62.
124. Brannan, K., et al., *mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription.* Molecular Cell, 2012. **46**(3): p. 311-324.
125. Beelman, C.A., et al., *An essential component of the decapping enzyme required for normal rates of mRNA turnover.* Nature, 1996. **382**(6592): p. 642-646.
126. LaGrandeur, T.E. and R. Parker, *Isolation and characterization of Dcp1p, the yeast mRNA decapping enzyme.* Embo Journal, 1998. **17**(5): p. 1487-1496.
127. Dunckley, T. and R. Parker, *The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif.* Embo Journal, 1999. **18**(19): p. 5411-5422.
128. Bitton, D.A., et al., *Widespread exon skipping triggers degradation by nuclear RNA surveillance in fission yeast.* Genome Research, 2015. **25**(6): p. 884-896.
129. Bousquet-Antonelli, C., C. Presutti, and D. Tollervey, *Identification of a regulated pathway for nuclear pre-mRNA turnover.* Cell, 2000. **102**(6): p. 765-775.
130. Gudipati, R.K., et al., *Extensive degradation of RNA precursors by the exosome in wild-type cells.* Mol Cell, 2012. **48**(3): p. 409-21.
131. Schneider, C., et al., *Transcriptome-wide Analysis of Exosome Targets.* Molecular Cell, 2012. **48**(3): p. 422-433.
132. Szczepinska, T., et al., *DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts.* Genome Research, 2015. **25**(11): p. 1622-1633.
133. Dziembowski, A., et al., *A single subunit, Dis3, is essentially responsible for yeast exosome core activity.* Nature Structural & Molecular Biology, 2007. **14**(1): p. 15-22.
134. Wasmuth, E.V. and C.D. Lima, *Exo- and Endoribonucleolytic Activities of Yeast Cytoplasmic and Nuclear RNA Exosomes Are Dependent on the Noncatalytic Core and Central Channel.* Molecular Cell, 2012. **48**(1): p. 133-144.
135. Vasiljeva, L., et al., *The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain.* Nat Struct Mol Biol, 2008. **15**(8): p. 795-804.
136. Gudipati, R.K., et al., *Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice.* Nature Structural & Molecular Biology, 2008. **15**(8): p. 786-794.
137. Steinmetz, E.J. and D.A. Brow, *Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear*

ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. Molecular and Cellular Biology, 1996. **16**(12): p. 6993-7003.
138. Conrad, N.K., et al., *A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II.* Genetics, 2000. **154**(2): p. 557-571.
139. Wilson, S.M., et al., *Characterization of nuclear polyadenylated RNA-binding proteins in Saccharomyces cerevisiae.* J Cell Biol, 1994. **127**(5): p. 1173-84.
140. Hobor, F., et al., *Recognition of Transcription Termination Signal by the Nuclear Polyadenylated RNA-binding (NAB) 3 Protein.* Journal of Biological Chemistry, 2011. **286**(5): p. 3645-3657.
141. Ursic, D., et al., *The yeast SEN1 gene is required for the processing of diverse RNA classes.* Nucleic Acids Res, 1997. **25**(23): p. 4778-85.
142. Steinmetz, E.J., et al., *Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase.* Molecular Cell, 2006. **24**(5): p. 735-746.
143. Ursic, D., D.J. DeMarini, and M.R. Culbertson, *Inactivation of the yeast Sen1 protein affects the localization of nucleolar proteins.* Molecular & General Genetics, 1995. **249**(6): p. 571-584.
144. Kim, H.D., J. Choe, and Y.S. Seo, *The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase.* Biochemistry, 1999. **38**(44): p. 14697-710.
145. Porrua, O. and D. Libri, *A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast.* Nat Struct Mol Biol, 2013. **20**(7): p. 884-91.
146. Martin-Tumasz, S. and D.A. Brow, *Saccharomyces cerevisiae Sen1 Helicase Domain Exhibits 5'- to 3'-Helicase Activity with a Preference for Translocation on DNA Rather than RNA.* J Biol Chem, 2015. **290**(38): p. 22880-9.
147. Kim, M., et al., *Distinct pathways for snoRNA and mRNA termination.* Molecular Cell, 2006. **24**(5): p. 723-734.
148. Arigo, J.T., et al., *Termination of cryptic unstable transcripts is directed by yeast RNA-Binding proteins Nrd1 and Nab3.* Molecular Cell, 2006. **23**(6): p. 841-851.
149. Creamer, T.J., et al., *Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1.* PLoS Genet, 2011. **7**(10): p. e1002329.
150. Kim, H., et al., *Gene-specific RNA polymerase II phosphorylation and the CTD code.* Nature Structural & Molecular Biology, 2010. **17**(10): p. 1279-+.
151. Mayer, A., et al., *CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II.* Science, 2012. **336**(6089): p. 1723-5.
152. LaCava, J., et al., *RNA degradation by the exosome is promoted by a nuclear polyade nylation complex.* Cell, 2005. **121**(5): p. 713-724.
153. Thiebaut, M., et al., *Transcription termination and nuclear degradation of cryptic unstable transcripts: A role for the Nrd1-Nab3 pathway in genome surveillance.* Molecular Cell, 2006. **23**(6): p. 853-864.
154. Jia, H.J., et al., *RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(19): p. 7292-7297.

155. Vasiljeva, L. and S. Buratowski, *Nrd1 interacts with the nuclear exosome for 3 ' processing of RNA polymerase II transcripts.* Molecular Cell, 2006. **21**(2): p. 239-248.
156. David, L., et al., *A high-resolution map of transcription in the yeast genome.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(14): p. 5320-5325.
157. Li, L., et al., *Genome-wide transcription analyses in rice using tiling microarrays.* Nature Genetics, 2006. **38**(1): p. 124-129.
158. Carninci, P., et al., *The transcriptional landscape of the mammalian genome.* Science, 2005. **309**(5740): p. 1559-1563.
159. Xu, Z., et al., *Bidirectional promoters generate pervasive transcription in yeast.* Nature, 2009. **457**(7232): p. 1033-7.
160. Neil, H., et al., *Widespread bidirectional promoters are the major source of cryptic transcripts in yeast.* Nature, 2009. **457**(7232): p. 1038-42.
161. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7934-8.
162. Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD.* Genes Dev, 2006. **20**(21): p. 2922-36.
163. West, M.L. and J.L. Corden, *Construction and Analysis of Yeast RNA Polymerase II Ctd Deletion and Substitution Mutations.* Genetics, 1995. **140**(4): p. 1223-1233.
164. Buratowski, S., *Progression through the RNA Polymerase II CTD Cycle.* Molecular Cell, 2009. **36**(4): p. 541-546.
165. Steinmetz, E.J., et al., *RNA-binding protein Nrd1 directs poly(A)-independent 3 '-end formation of RNA polymerase II transcripts.* Nature, 2001. **413**(6853): p. 327-331.
166. Suh, H., et al., *Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II.* Mol Cell, 2016. **61**(2): p. 297-304.
167. Trigon, S., et al., *Characterization of the residues phosphorylated in vitro by different C-terminal domain kinases.* J Biol Chem, 1998. **273**(12): p. 6769-75.
168. Schuller, R., et al., *Heptad-Specific Phosphorylation of RNA Polymerase II CTD.* Mol Cell, 2016. **61**(2): p. 305-14.
169. Zhang, J. and J.L. Corden, *Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit.* J Biol Chem, 1991. **266**(4): p. 2290-6.
170. Corden, J.L., *RNA polymerase II C-terminal domain: Tethering transcription to transcript and template.* Chem Rev, 2013. **113**(11): p. 8423-55.
171. Stiller, J.W. and M.S. Cook, *Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs.* Eukaryot Cell, 2004. **3**(3): p. 735-40.
172. Egloff, S. and S. Murphy, *Cracking the RNA polymerase II CTD code.* Trends Genet, 2008. **24**(6): p. 280-8.
173. Zhang, M., et al., *Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code.* ACS Chem Biol, 2012. **7**(8): p. 1462-70.
174. Meinhart, A. and P. Cramer, *Recognition of RNA polymerase II carboxy-terminal domain by 3 '-RNA-processing factors.* Nature, 2004. **430**(6996): p. 223-226.

175. Noble, C.G., et al., *Key features of the interaction between Pcf11 CID and RNA polymerase IICTD.* Nature Structural & Molecular Biology, 2005. **12**(2): p. 144-151.
176. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom ngstrom resolution.* Science, 2001. **292**(5523): p. 1863-1876.
177. Harding, M.M., *NMR studies on YSPTSPSY: implications for the design of DNA bisintercalators.* J Med Chem, 1992. **35**(25): p. 4658-64.
178. Zhang, J. and J.L. Corden, *Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit.* J Biol Chem, 1991. **266**(4): p. 2297-302.
179. Portz, B., et al., *Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain.* Nature Communications, 2017. **8**.
180. Komarnitsky, P., E.J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.* Genes & Development, 2000. **14**(19): p. 2452-2460.
181. Heidemann, M., et al., *Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription.* Biochim Biophys Acta, 2013. **1829**(1): p. 55-62.
182. Serizawa, H., J.W. Conaway, and R.C. Conaway, *Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription.* Nature, 1993. **363**(6427): p. 371-4.
183. Mayer, A., et al., *Uniform transitions of the general RNA polymerase II transcription complex.* Nature Structural & Molecular Biology, 2010. **17**(10): p. 1272-+.
184. Bataille, A.R., et al., *A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes.* Molecular Cell, 2012. **45**(2): p. 158-170.
185. Zhang, D.W., et al., *Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination.* J Biol Chem, 2012. **287**(11): p. 8541-51.
186. Baskaran, R., S.R. Escobar, and J.Y.J. Wang, *Nuclear c-Abl is a COOH-terminal repeated domain (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: Possible role in transcription elongation.* Cell Growth & Differentiation, 1999. **10**(6): p. 387-396.
187. Schreieck, A., et al., *RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7.* Nature Structural & Molecular Biology, 2014. **21**(2): p. 175-+.
188. Sun, M., et al., *A Tandem SH2 Domain in Transcription Elongation Factor Spt6 Binds the Phosphorylated RNA Polymerase II C-terminal Repeat Domain (CTD).* Journal of Biological Chemistry, 2010. **285**(53): p. 41597-41603.
189. Cho, E.J., et al., *Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain.* Genes Dev, 2001. **15**(24): p. 3319-29.
190. Wood, A. and A. Shilatifard, *Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb.* Cell Cycle, 2006. **5**(10): p. 1066-8.
191. Yoh, S.M., et al., *The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export.* Genes Dev, 2007. **21**(2): p. 160-74.

192. David, C.J., et al., *The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex.* Genes Dev, 2011. **25**(9): p. 972-83.
193. MacKellar, A.L. and A.L. Greenleaf, *Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II.* J Biol Chem, 2011. **286**(42): p. 36385-95.
194. Kizer, K.O., et al., *A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation.* Mol Cell Biol, 2005. **25**(8): p. 3305-16.
195. Vojnic, E., et al., *Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription.* J Biol Chem, 2006. **281**(1): p. 13-5.
196. Fuchs, S.M., et al., *RNA polymerase II carboxyl-terminal domain phosphorylation regulates protein stability of the Set2 methyltransferase and histone H3 di- and trimethylation at lysine 36.* J Biol Chem, 2012. **287**(5): p. 3249-56.
197. Pirngruber, J., et al., *CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing.* EMBO Rep, 2009. **10**(8): p. 894-900.
198. Hintermair, C., et al., *Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation.* EMBO J, 2012. **31**(12): p. 2784-97.
199. Schwer, B. and S. Shuman, *Deciphering the RNA Polymerase II CTD Code in Fission Yeast.* Molecular Cell, 2011. **43**(2): p. 311-318.
200. Hsin, J.P., A. Sheth, and J.L. Manley, *RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3 ' End Processing.* Science, 2011. **334**(6056): p. 683-686.
201. Krishnamurthy, S., et al., *Ssu72 is an RNA polymerase IICTD phosphatase.* Molecular Cell, 2004. **14**(3): p. 387-394.
202. Guidi, B.W., et al., *Mutual targeting of mediator and the TFIIH kinase Kin28.* Journal of Biological Chemistry, 2004. **279**(28): p. 29114-29120.
203. Mosley, A.L., et al., *Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation.* Mol Cell, 2009. **34**(2): p. 168-78.
204. Boeing, S., et al., *RNA Polymerase II C-terminal Heptarepeat Domain Ser-7 Phosphorylation Is Established in a Mediator-dependent Fashion.* Journal of Biological Chemistry, 2010. **285**(1): p. 188-196.
205. Irani, S., et al., *Structure of Saccharomyces cerevisiae Rtr1 reveals an active site for an atypical phosphatase.* Sci Signal, 2016. **9**(417): p. ra24.
206. Hsu, P.L., et al., *Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD.* J Mol Biol, 2014. **426**(16): p. 2970-81.
207. Gibney, P.A., et al., *Rtr1 is the Saccharomyces cerevisiae homolog of a novel family of RNA polymerase II-binding proteins.* Eukaryot Cell, 2008. **7**(6): p. 938-48.
208. Sogaard, T.M. and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator.* J Biol Chem, 2007. **282**(19): p. 14113-20.

209. Jiang, Y., M. Yan, and J.D. Gralla, *A three-step pathway of transcription initiation leading to promoter clearance at an activation RNA polymerase II promoter.* Mol Cell Biol, 1996. **16**(4): p. 1614-21.
210. Liu, Y., et al., *Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex.* Molecular and Cellular Biology, 2004. **24**(4): p. 1721-1735.
211. Cho, E.J., et al., *mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain.* Genes & Development, 1997. **11**(24): p. 3319-3326.
212. Tietjen, J.R., et al., *Chemical-genomic dissection of the CTD code.* Nat Struct Mol Biol, 2010. **17**(9): p. 1154-61.
213. Brookes, E., et al., *Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs.* Cell Stem Cell, 2012. **10**(2): p. 157-170.
214. Kim, T.S., et al., *RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast.* Genome Biology, 2010. **11**(7).
215. Egloff, S., et al., *Serine-7 of the RNA polymerase IICTD is specifically required for snRNA gene expression.* Science, 2007. **318**(5857): p. 1777-1779.
216. Xiang, K.H., et al., *Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex.* Nature, 2010. **467**(7316): p. 729-U124.
217. Werner-Allen, J.W., et al., *cis-Proline-mediated Ser(P)(5) Dephosphorylation by the RNA Polymerase II C-terminal Domain Phosphatase Ssu72.* Journal of Biological Chemistry, 2011. **286**(7): p. 5717-5726.
218. Wada, T., et al., *DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs.* Genes & Development, 1998. **12**(3): p. 343-356.
219. Bernecky, C., J.M. Plitzko, and P. Cramer, *Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp.* Nature Structural & Molecular Biology, 2017. **24**(10): p. 809-+.
220. Ehara, H., et al., *Structure of the complete elongation complex of RNA polymerase II with basal factors.* Science, 2017. **357**(6354): p. 921-924.
221. Yamada, T., et al., *P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation.* Molecular Cell, 2006. **21**(2): p. 227-237.
222. Sanso, M., et al., *P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates.* Genes Dev, 2016. **30**(1): p. 117-31.
223. Klreeva, M., et al., *RNA-DNA and DNA-DNA base-pairing at the upstream edge of the transcription bubble regulate translocation of RNA polymerase and transcription rate.* Nucleic Acids Research, 2018. **46**(11): p. 5764-5775.
224. Van Oss, S.B., C.E. Cucinotta, and K.M. Arndt, *Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation.* Trends in Biochemical Sciences, 2017. **42**(10): p. 788-798.
225. Ardehali, M.B., et al., *Spt6 enhances the elongation rate of RNA polymerase II in vivo.* EMBO J, 2009. **28**(8): p. 1067-77.

226. Endoh, M., et al., *Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro.* Molecular and Cellular Biology, 2004. **24**(8): p. 3324-3336.
227. Kaplan, C.D., L. Laprade, and F. Winston, *Transcription elongation factors repress transcription initiation from cryptic sites.* Science, 2003. **301**(5636): p. 1096-1099.
228. Sdano, M.A., et al., *A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription.* Elife, 2017. **6**.
229. Yu, M., et al., *RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II.* Science, 2015. **350**(6266): p. 1383-1386.
230. Shi, X.M., et al., *Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme.* Molecular and Cellular Biology, 1997. **17**(3): p. 1160-1169.
231. Mueller, C.L., et al., *The Paf1 complex has functions independent of actively transcribing RNA polymerase II.* Molecular Cell, 2004. **14**(4): p. 447-456.
232. Amrich, C.G., et al., *Cdc73 Subunit of Paf1 Complex Contains C-terminal Ras-like Domain That Promotes Association of Paf1 Complex with Chromatin.* Journal of Biological Chemistry, 2012. **287**(14): p. 10863-10875.
233. Qiu, H., et al., *Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex.* EMBO J, 2012. **31**(16): p. 3494-505.
234. Wier, A.D., et al., *Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(43): p. 17290-17295.
235. Mayekar, M.K., R.G. Gardner, and K.M. Arndt, *The Recruitment of the Saccharomyces cerevisiae Paf1 Complex to Active Genes Requires a Domain of Rtf1 That Directly Interacts with the Spt4-Spt5 Complex.* Molecular and Cellular Biology, 2013. **33**(16): p. 3259-3273.
236. Wu, L., et al., *H2B ubiquitylation promotes RNA Pol II processivity via PAF1 and pTEFb.* Mol Cell, 2014. **54**(6): p. 920-931.
237. Porrua, O. and D. Libri, *Transcription termination and the control of the transcriptome: why, where and how to stop.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 190-202.
238. Colin, J., et al., *Roadblock termination by reb1p restricts cryptic and readthrough transcription.* Mol Cell, 2014. **56**(5): p. 667-80.
239. Matsumoto, Y., et al., *Autogenous regulation of the gene for transcription termination factor rho in Escherichia coli: localization and function of its attenuators.* J Bacteriol, 1986. **166**(3): p. 945-58.
240. Gudipati, R.K., et al., *The yeast RPL9B gene is regulated by modulation between two modes of transcription termination.* EMBO J, 2012. **31**(10): p. 2427-37.
241. Thiebaut, M., et al., *Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae.* Mol Cell, 2008. **31**(5): p. 671-82.

242. Kuehner, J.N. and D.A. Brow, *Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation.* Mol Cell, 2008. **31**(2): p. 201-11.
243. Callahan, K.P. and J.S. Butler, *TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6.* J Biol Chem, 2010. **285**(6): p. 3540-7.
244. Jensen, T.H., A. Jacquier, and D. Libri, *Dealing with Pervasive Transcription.* Molecular Cell, 2013. **52**(4): p. 473-484.
245. Shearwin, K.E., B.P. Callen, and J.B. Egan, *Transcriptional interference--a crash course.* Trends Genet, 2005. **21**(6): p. 339-45.
246. Castelnuovo, M., et al., *Bimodal expression of PHO84 is modulated by early termination of antisense transcription.* Nat Struct Mol Biol, 2013. **20**(7): p. 851-8.
247. Martens, J.A., P.Y.J. Wu, and F. Winston, *Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae.* Genes & Development, 2005. **19**(22): p. 2695-2704.
248. Fox, M.J., et al., *The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway.* PLoS Genet, 2015. **11**(2): p. e1004999.
249. Grenier St-Sauveur, V., et al., *Poly(A) tail-mediated gene regulation by opposing roles of Nab2 and Pab2 nuclear poly(A)-binding proteins in pre-mRNA decay.* Mol Cell Biol, 2013. **33**(23): p. 4718-31.
250. Garland, W., et al., *Rrp47 functions in RNA surveillance and stable RNA processing when divorced from the exoribonuclease and exosome-binding domains of Rrp6.* Rna-a Publication of the Rna Society, 2013. **19**(12): p. 1659-1668.
251. Mitchell, P., et al., *The exosome: A conserved eukaryotic RNA processing complex containing multiple 3'->5' exoribonucleases.* Cell, 1997. **91**(4): p. 457-466.
252. Allmang, C., et al., *Functions of the exosome in rRNA, snoRNA and snRNA synthesis.* Embo Journal, 1999. **18**(19): p. 5399-5410.
253. Rondón, A.G., et al., *Fail-Safe Transcriptional Termination for Protein-Coding Genes in S. cerevisiae.* Molecular Cell, 2009. **36**(1): p. 88-98.
254. Castelnuovo, M., et al., *Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast.* Nucleic Acids Res, 2014. **42**(7): p. 4348-62.
255. Ozsolak, F., et al., *Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation.* Cell, 2010. **143**(6): p. 1018-1029.
256. Nedea, E., et al., *Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3 '-ends.* Journal of Biological Chemistry, 2003. **278**(35): p. 33000-33010.
257. Kuehner, J.N., E.L. Pearson, and C. Moore, *Unravelling the means to an end: RNA polymerase II transcription termination.* Nature Reviews Molecular Cell Biology, 2011. **12**(5): p. 283-294.

258. Xiang, K., L. Tong, and J.L. Manley, *Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery.* Mol Cell Biol, 2014. **34**(11): p. 1894-910.
259. Mischo, H.E. and N.J. Proudfoot, *Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast.* Biochim Biophys Acta, 2013. **1829**(1): p. 174-85.
260. Kyburz, A., et al., *The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation.* Nucleic Acids Res, 2003. **31**(14): p. 3936-45.
261. Schreieck, A., et al., *RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7.* Nat Struct Mol Biol, 2014. **21**(2): p. 175-179.
262. Yoshio, T.D. and J.L. Manley, *RNA recognition by the human polyadenylation factor CstF.* Molecular and Cellular Biology, 1997. **17**(7): p. 3907-3914.
263. Murthy, K.G.K. and J.L. Manley, *The 160-Kd Subunit of Human Cleavage Polyadenylation Specificity Factor Coordinates Pre-Messenger-Rna 3'-End Formation.* Genes & Development, 1995. **9**(21): p. 2672-2683.
264. Nag, A., K. Narsinh, and H.G. Martinson, *The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase.* Nature Structural & Molecular Biology, 2007. **14**(7): p. 662-669.
265. Barabino, S.M., M. Ohnacker, and W. Keller, *Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs.* EMBO J, 2000. **19**(14): p. 3778-87.
266. Barabino, S.M., et al., *The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins.* Genes Dev, 1997. **11**(13): p. 1703-16.
267. Kaufmann, I., et al., *Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase.* EMBO J, 2004. **23**(3): p. 616-26.
268. Mandel, C.R., et al., *Polyadenylation factor CPSF-73 is the pre-mRNA 3 '-end-processing endonuclease.* Nature, 2006. **444**(7121): p. 953-956.
269. Zhelkovsky, A., et al., *The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3 '-end processing in Saccharomyces cerevisiae.* Rna, 2006. **12**(3): p. 435-445.
270. Wahle, E., *Purification and Characterization of a Mammalian Polyadenylate Polymerase Involved in the 3' End Processing of Messenger-Rna Precursors.* Journal of Biological Chemistry, 1991. **266**(5): p. 3131-3139.
271. Bienroth, S., et al., *Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors.* J Biol Chem, 1991. **266**(29): p. 19768-76.
272. Carroll, K.L., et al., *Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements.* Rna-a Publication of the Rna Society, 2007. **13**(3): p. 361-373.
273. Nedea, E., et al., *The glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes.* Molecular Cell, 2008. **29**(5): p. 577-587.

274. Porrua, O., et al., *In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination.* Embo Journal, 2012. **31**(19): p. 3935-3948.
275. Preker, P.J., et al., *A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor.* EMBO J, 1997. **16**(15): p. 4727-37.
276. Ganem, C., et al., *Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast.* Embo Journal, 2003. **22**(7): p. 1588-1598.
277. Steinmetz, E.J. and D.A. Brow, *Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription.* Molecular and Cellular Biology, 2003. **23**(18): p. 6339-6349.
278. Cheng, H.L., X.Y. He, and C. Moore, *The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3.* Molecular and Cellular Biology, 2004. **24**(7): p. 2932-2943.
279. Roguev, A., et al., *The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4.* EMBO J, 2001. **20**(24): p. 7137-48.
280. Russnak, R., K.W. Nehrke, and T. Platt, *Ref2 Encodes an Rna-Binding Protein Directly Involved in Yeast Messenger-Rna 3'-End Formation.* Molecular and Cellular Biology, 1995. **15**(3): p. 1689-1697.
281. Dheur, S., et al., *Pti1p and Ref2p found in association with the mRNA 3 ' end formation complex direct snoRNA maturation.* Embo Journal, 2003. **22**(11): p. 2831-2840.
282. Lidschreiber, M., et al., *The APT complex is involved in non-coding RNA transcription and is distinct from CPF.* Nucleic Acids Res, 2018. **46**(21): p. 11528-11538.
283. Hernandez, H. and C.V. Robinson, *Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry.* Nature Protocols, 2007. **2**(3): p. 715-726.
284. Casanal, A., et al., *Architecture of eukaryotic mRNA 3 '-end processing machinery.* Science, 2017. **358**(6366): p. 1056-1059.
285. Chambers, R.S. and M.E. Dahmus, *Purification and Characterization of a Phosphatase from Hela-Cells Which Dephosphorylates the C-Terminal Domain of Rna-Polymerase-Ii.* Journal of Biological Chemistry, 1994. **269**(42): p. 26243-26248.
286. Kimura, M., H. Suzuki, and A. Ishihama, *Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II.* Molecular and Cellular Biology, 2002. **22**(5): p. 1577-1588.
287. Archambault, J., et al., *An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(26): p. 14300-14305.

288. Hausmann, S., et al., *Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases.* Journal of Biological Chemistry, 2005. **280**(45): p. 37681-37688.
289. Ghosh, A., S. Shuman, and C.D. Lima, *The Structure of Fcp1, an Essential RNA Polymerase II CTD Phosphatase.* Molecular Cell, 2008. **32**(4): p. 478-490.
290. Wostenberg, C., et al., *Atomistic simulations reveal structural disorder in the RAP74-FCP1 complex.* J Phys Chem B, 2011. **115**(46): p. 13731-9.
291. Kobor, M.S., et al., *A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 2000. **20**(20): p. 7438-7449.
292. Chambers, R.S., et al., *The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB.* J Biol Chem, 1995. **270**(25): p. 14962-9.
293. Suh, M.H., et al., *Fcp1 directly recognizes the C-terminal domain (CTD) and interacts with a site on RNA polymerase II distinct from the CTD.* Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17314-9.
294. Fuda, N.J., et al., *Fcp1 Dephosphorylation of the RNA Polymerase II C-Terminal Domain Is Required for Efficient Transcription of Heat Shock Genes.* Molecular and Cellular Biology, 2012. **32**(17): p. 3428-3437.
295. Kobor, M.S., et al., *An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae.* Mol Cell, 1999. **4**(1): p. 55-62.
296. Sun, Z.W. and M. Hampsey, *Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo.* Molecular and Cellular Biology, 1996. **16**(4): p. 1557-1566.
297. Dichtl, B., et al., *A role for SSU72 in balancing RNA polymerase II transcription elongation and termination.* Molecular Cell, 2002. **10**(5): p. 1139-1150.
298. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. **415**(6868): p. 141-147.
299. Pappas, D.L. and M. Hampsey, *Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 2000. **20**(22): p. 8343-8351.
300. Hampsey, M., et al., *Control of eukaryotic gene expression: Gene loops and transcriptional memory.* Advances in Enzyme Regulation, Vol 51, 2011. **51**: p. 118-125.
301. Jeronimo, C., et al., *Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme.* Mol Cell, 2007. **27**(2): p. 262-74.
302. Hunter, G.O., et al., *Phosphatase Rtr1 Regulates Global Levels of Serine 5 RNA Polymerase II C-Terminal Domain Phosphorylation and Cotranscriptional Histone Methylation.* Mol Cell Biol, 2016. **36**(17): p. 2236-45.
303. Krishnamurthy, S., et al., *Functional Interaction of the Ess1 Prolyl Isomerase with Components of the RNA Polymerase II Initiation and Termination Machineries.* Molecular and Cellular Biology, 2009. **29**(11): p. 2925-2934.

304. Singh, N., et al., *The Ess1 Prolyl Isomerase Is Required for Transcription Termination of Small Noncoding RNAs via the Nrd1 Pathway.* Molecular Cell, 2009. **36**(2): p. 255-266.
305. Mayfield, J.E., et al., *Chemical Tools To Decipher Regulation of Phosphatases by Proline Isomerization on Eukaryotic RNA Polymerase II.* Acs Chemical Biology, 2015. **10**(10): p. 2405-2414.
306. Florens, L. and M.P. Washburn, *Proteomic analysis by multidimensional protein identification technology.* Methods Mol Biol, 2006. **328**: p. 159-75.
307. Kawamoto, S., et al., *Expression profiles of active genes in human and mouse livers.* Gene, 1996. **174**(1): p. 151-158.
308. Anderson, L. and J. Seilhamer, *A comparison of selected mRNA and protein abundances in human liver.* Electrophoresis, 1997. **18**(3-4): p. 533-537.
309. Futcher, B., et al., *A sampling of the yeast proteome.* Molecular and Cellular Biology, 1999. **19**(11): p. 7357-7368.
310. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Molecular and Cellular Biology, 1999. **19**(3): p. 1720-1730.
311. Das, R., et al., *SR proteins function in coupling RNAP II transcription to pre-mRNA splicing.* Mol Cell, 2007. **26**(6): p. 867-81.
312. Fountoulakis, M., et al., *Enrichment of low abundance proteins of Escherichia coli by hydroxyapatite chromatography.* Electrophoresis, 1999. **20**(11): p. 2181-95.
313. Fountoulakis, M., M.F. Takacs, and B. Takacs, *Enrichment of low-copy-number gene products by hydrophobic interaction chromatography.* J Chromatogr A, 1999. **833**(2): p. 157-68.
314. Santoni, V., M. Molloy, and T. Rabilloud, *Membrane proteins and proteomics: un amour impossible?* Electrophoresis, 2000. **21**(6): p. 1054-70.
315. Corthals, G.L., et al., *The dynamic range of protein expression: a challenge for proteomic research.* Electrophoresis, 2000. **21**(6): p. 1104-15.
316. Oh-Ishi, M., M. Satoh, and T. Maeda, *Preparative two-dimensional gel electrophoresis with agarose gels in the first dimension for high molecular mass proteins.* Electrophoresis, 2000. **21**(9): p. 1653-69.
317. Washburn, M.P., D. Wolters, and J.R. Yates, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nature Biotechnology, 2001. **19**(3): p. 242-247.
318. Florens, L., et al., *A proteomic view of the Plasmodium falciparum life cycle.* Nature, 2002. **419**(6906): p. 520-526.
319. Koller, A., et al., *Proteomic survey of metabolic pathways in rice.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11969-74.
320. Pan, Y., et al., *Identification of biochemical adaptations in hyper- or hypocontractile hearts from phospholamban mutant mice by expression proteomics.* Proc Natl Acad Sci U S A, 2004. **101**(8): p. 2241-6.
321. Washburn, M.P., et al., *Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3107-12.
322. Wu, C.C., et al., *A method for the comprehensive proteomic analysis of membrane proteins.* Nat Biotechnol, 2003. **21**(5): p. 532-8.

323. Graumann, J., et al., *Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast.* Molecular & Cellular Proteomics, 2004. **3**(3): p. 226-237.
324. Link, A.J., et al., *Direct analysis of protein complexes using mass spectrometry.* Nature Biotechnology, 1999. **17**(7): p. 676-682.
325. Fenn, J.B., et al., *Electrospray Ionization for Mass-Spectrometry of Large Biomolecules.* Science, 1989. **246**(4926): p. 64-71.
326. Griffin, P.R., et al., *Structural-Analysis of Proteins by Capillary Hplc Electrospray Tandem Mass-Spectrometry.* International Journal of Mass Spectrometry and Ion Processes, 1991. **111**: p. 131-149.
327. Emmett, M.R. and R.M. Caprioli, *Micro-Electrospray Mass-Spectrometry - Ultra-High-Sensitivity Analysis of Peptides and Proteins.* Journal of the American Society for Mass Spectrometry, 1994. **5**(7): p. 605-613.
328. Olsen, J.V., et al., *Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis.* Sci Signal, 2010. **3**(104): p. ra3.
329. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-33.
330. Gross, J.H., *Mass spectrometry : a textbook*. 2004, Springer: Berlin ; New York. p. xviii, 518 pages.
331. De Hoffmann, E. and V. Stroobant, *Mass spectrometry : principles and applications*. 3rd ed. 2007, Chichester: John Wiley. xii, 489 p.
332. Eng, J.K., A.L. Mccormack, and J.R. Yates, *An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database.* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.
333. Krogan, N.J., et al., *Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II.* Molecular and Cellular Biology, 2003. **23**(12): p. 4207-4218.
334. Xiao, T., et al., *Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast.* Genes Dev, 2003. **17**(5): p. 654-63.
335. Li, B., et al., *The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II.* Journal of Biological Chemistry, 2003. **278**(11): p. 8897-8903.
336. Li, J., D. Moazed, and S.P. Gygi, *Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation.* J Biol Chem, 2002. **277**(51): p. 49383-8.
337. Walian, P., T.A. Cross, and B.K. Jap, *Structural genomics of membrane proteins.* Genome Biology, 2004. **5**(4).
338. Sinz, A., *Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions.* Mass Spectrom Rev, 2006. **25**(4): p. 663-82.
339. Sinz, A., *The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks.* Expert Rev Proteomics, 2014. **11**(6): p. 733-43.

340. Back, J.W., et al., *Chemical cross-linking and mass spectrometry for protein structural modeling.* J Mol Biol, 2003. **331**(2): p. 303-13.
341. Sinz, A., *Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes.* Journal of Mass Spectrometry, 2003. **38**(12): p. 1225-1237.
342. Dihazi, G.H. and A. Sinz, *Mapping low-resolution three-dimensional protein structures using chemical cross-linking and Fourier transform ion-cyclotron resonance mass spectrometry.* Rapid Communications in Mass Spectrometry, 2003. **17**(17): p. 2005-2014.
343. Sinz, A., et al., *Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology.* Protein Science, 2015. **24**(8): p. 1193-1209.
344. Mayne, S.L.N. and H.G. Patterton, *Bioinformatics tools for the structural elucidation of multi-subunit protein complexes by mass spectrometric analysis of protein-protein cross-links.* Briefings in Bioinformatics, 2011. **12**(6): p. 660-671.
345. Staros, J.V., *N-hydroxysulfosuccinimide active esters: bis(N-hydroxysulfosuccinimide) esters of two dicarboxylic acids are hydrophilic, membrane-impermeant, protein cross-linkers.* Biochemistry, 1982. **21**(17): p. 3950-5.
346. Kao, A.H., et al., *Development of a Novel Cross-linking Strategy for Fast and Accurate Identification of Cross-linked Peptides of Protein Complexes.* Molecular & Cellular Proteomics, 2011. **10**(1).
347. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* Science, 2007. **316**(5830): p. 1484-1488.
348. Madler, S., et al., *Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities.* Journal of Mass Spectrometry, 2009. **44**(5): p. 694-706.
349. Reyes-Reyes, M. and M. Hampsey, *Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation.* Mol Cell Biol, 2007. **27**(3): p. 926-36.
350. Loya, T.J., T.W. O'Rourke, and D. Reines, *A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3.* Nucleic Acids Res, 2012. **40**(15): p. 7476-91.
351. Jenks, M.H., T.W. O'Rourke, and D. Reines, *Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast.* Mol Cell Biol, 2008. **28**(12): p. 3883-93.
352. Rhee, H.S. and B.F. Pugh, *Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution.* Cell, 2011. **147**(6): p. 1408-19.
353. Choi, H., et al., *SAINT: probabilistic scoring of affinity purification-mass spectrometry data.* Nature Methods, 2011. **8**(1): p. 70-U100.
354. Winzeler, E.A., et al., *Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis.* Science, 1999. **285**(5429): p. 901-6.
355. Mosley, A.L., et al., *Highly reproducible label free quantitative proteomic analysis of RNA polymerase complexes.* Mol Cell Proteomics, 2011. **10**(2): p. M110 000687.

356. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nature biotechnology, 2001. **19**(3): p. 242-7.
357. Choi, H., et al., *SAINT: probabilistic scoring of affinity purification-mass spectrometry data.* Nat Methods, 2011. **8**(1): p. 70-3.
358. Teo, G., et al., *SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software.* J Proteomics, 2014. **100**: p. 37-43.
359. Smith-Kinnaman, W.R., et al., *The interactome of the atypical phosphatase Rtr1 in Saccharomyces cerevisiae.* Mol Biosyst, 2014. **10**(7): p. 1730-41.
360. Mellacheruvu, D., et al., *The CRAPome: a contaminant repository for affinity purification-mass spectrometry data.* Nat Methods, 2013. **10**(8): p. 730-6.
361. Knight, J.D.R., et al., *ProHits-viz: a suite of web tools for visualizing interaction proteomics data.* Nat Methods, 2017. **14**(7): p. 645-646.
362. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-140.
363. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
364. Wal, M. and B.F. Pugh, *Genome-wide mapping of nucleosome positions in yeast using high-resolution MNase ChIP-Seq.* Methods Enzymol, 2012. **513**: p. 233-50.
365. Shen, L., et al., *ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases.* BMC Genomics, 2014. **15**: p. 284.
366. Breese, M.R. and Y. Liu, *NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets.* Bioinformatics, 2013. **29**(4): p. 494-6.
367. Gudipati, R.K., et al., *Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice.* Nat Struct Mol Biol, 2008. **15**(8): p. 786-94.
368. Vasiljeva, L., et al., *The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain.* Nature Structural & Molecular Biology, 2008. **15**(8): p. 795-804.
369. Arndt, K.M. and D. Reines, *Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II.* Annu Rev Biochem, 2015. **84**: p. 381-404.
370. Casanal, A., et al., *Architecture of eukaryotic mRNA 3'-end processing machinery.* Science, 2017. **358**(6366): p. 1056-1059.
371. Lunde, B.M., et al., *Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain.* Nat Struct Mol Biol, 2010. **17**(10): p. 1195-201.
372. Bedard, L.G., et al., *Quantitative Analysis of Dynamic Protein Interactions during Transcription Reveals a Role for Casein Kinase II in Polymerase-associated Factor (PAF) Complex Phosphorylation and Regulation of Histone H2B Monoubiquitylation.* J Biol Chem, 2016. **291**(26): p. 13410-20.

373. Hunter, G.O., et al., *The phosphatase Rtr1 regulates global levels of serine 5 RNA Polymerase II C-terminal domain phosphorylation and cotranscriptional histone methylation.* Mol Cell Biol, 2016.
374. Vasiljeva, L. and S. Buratowski, *Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts.* Mol Cell, 2006. **21**(2): p. 239-48.
375. Jiang, L., et al., *Synthetic spike-in standards for RNA-seq experiments.* Genome Res, 2011. **21**(9): p. 1543-51.
376. Schulz, D., et al., *Transcriptome surveillance by selective termination of noncoding RNA synthesis.* Cell, 2013. **155**(5): p. 1075-87.
377. Fox, M.J., et al., *The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway.* PLoS genetics, 2015. **10**(2): p. e1004999.
378. Kopcewicz, K.A., T.W. O'Rourke, and D. Reines, *Metabolic regulation of IMD2 transcription and an unusual DNA element that generates short transcripts.* Mol Cell Biol, 2007. **27**(8): p. 2821-9.
379. Merran, J. and J.L. Corden, *Yeast RNA-Binding Protein Nab3 Regulates Genes Involved in Nitrogen Metabolism.* Mol Cell Biol, 2017. **37**(18).
380. Steinmetz, E.J., et al., *cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II.* Mol Cell Biol, 2006. **26**(7): p. 2688-96.
381. Hazelbaker, D.Z., et al., *Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination.* Mol Cell, 2013. **49**(1): p. 55-66.
382. Vasiljeva, L., et al., *Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin.* Mol Cell, 2008. **29**(3): p. 313-23.
383. Arigo, J.T., et al., *Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3.* Mol Cell, 2006. **23**(6): p. 841-51.
384. Fasken, M.B., R.N. Laribee, and A.H. Corbett, *Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1.* PLoS Genet, 2015. **11**(3): p. e1005044.
385. Lemay, J.F., et al., *The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs.* Mol Cell, 2010. **37**(1): p. 34-45.
386. LaCava, J., et al., *RNA degradation by the exosome is promoted by a nuclear polyadenylation complex.* Cell, 2005. **121**(5): p. 713-24.
387. Wickham, H., *ggplot2 - Elegant Graphics for Data Analysis*. 2016: Springer-Verlag, New York. 260.
388. He, X., et al., *Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1.* Genes Dev, 2003. **17**(8): p. 1030-42.
389. Wani, S., et al., *Vertebrate Ssu72 regulates and coordinates 3'-end formation of RNAs transcribed by RNA polymerase II.* PLoS One, 2014. **9**(8): p. e106040.
390. van Nues, R., et al., *Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress.* Nat Commun, 2017. **8**(1): p. 12.
391. Ganem C., D.F., Torchet C., Jacq C., Quevillon-Cheruel S., Labesse G., Facca C., Faye G., *Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast.* EMBO J, 2003. **22**: p. 1588-1598.

392. Guan, K., et al., *A yeast protein phosphatase related to the vaccinia virus VH1 phosphatase is induced by nitrogen starvation.* Proc Natl Acad Sci U S A, 1992. **89**(24): p. 12175-9.
393. Schroeder, S.C., et al., *Dynamic association of capping enzymes with transcribing RNA polymerase II.* Genes & Development, 2000. **14**(19): p. 2435-2440.
394. Bohm, S. and A. Buchberger, *The budding yeast Cdc48(Shp1) complex promotes cell cycle progression by positive regulation of protein phosphatase 1 (Glc7).* PLoS One, 2013. **8**(2): p. e56486.
395. Ni, Z.Y., et al., *RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation.* Nature Structural & Molecular Biology, 2014. **21**(8): p. 686-695.
396. Egloff, S., et al., *Ser7 Phosphorylation of the CTD Recruits the RPAP2 Ser5 Phosphatase to snRNA Genes.* Molecular Cell, 2012. **45**(1): p. 111-122.

CURRICULUM VITAE

Jose Fabian Victorino

Education

B.S. Biology, September 2010, University of California Irvine, Irvine, CA

M.S. Biological Sciences, July 2014, CSU Dominguez Hills, Carson, CA

Ph.D. Biochemistry and Molecular Biology, May 2020 Indiana University, Indianapolis, IN

Research Experience

Proteomics-Genomics, Graduate Student, IU School of Medicine (5/2015-01/2020)
Duties include: Culture and purify proteins and their complexes involved in transcription termination from *S. cerevisiae* to analyze protein-protein interactions using Mass Spectrometry using bioinformatics tools. Culture and purify RNA from *S. cerevisiae* with different mutations in proteins known to affect transcription and gage the effects of the deletions using real-time qPCR, ChIP-Seq, ChIP-Exo, Northern Blot, Cloning, Site directed mutagenesis etc.

Urology, Graduate Student, LA BioMed, Torrance, CA (1/2013-5/2014)
Duties include: Culture and maintain several types of cells including penis derived stem cells, induced pluripotent stem cells and muscle derived stem cells. Western blotting, immunohistochemistry, and drug testing for the purpose of studying the causes and possible treatments of erectile dysfunction.

Ant Ecology, Research Internship, La Selva Biological Station, (8/1/2013-8/14/2013)
Puerto Viejo de Sarapiqui, Costa Rica
Duties include: Excavating, collecting, organizing and identifying different types of ants within entire colonies of *Ectatomma ruidum.* Help analyze data to study the relationship between intraspecific cleptobiosis and productivity of social insects.

Cytogenetics, Graduate Student, (5/2012-8/2013)
IU School of Medicine, Indianapolis, IN
Duties include: Culture and maintain tumor and normal human cell lines; breast epithelial cells, and pancreatic cells. Perform FISH, Anaphase arrest, IC_{50}, Giemsa staining and metaphase arrested assays to analyze the effects of drugs targeting mitotic checkpoint component on normal and cancer cell lines specifically looking at the number of numerical chromosomal abnormalities as well as cell death and the ability of a certain drug to decrease TNBC cell viability.

Genetics: Infectious Diseases, Research Internship, (10/2011-2/2012)
LABioMed, Torrance, CA
Duties included: Culture and maintain different bacterial cell lines *of Staphylococcus aureus*. Perform kill curve, cytochrome c, 96 well plate, peptide susceptibility, and population assays to analyze the effects of different gene deletions in the model organism's growth rate, death rate, response to different human peptides, and population densities.

Genome Evolution Research Assistant, UCI, Irvine, CA (6/2010-10/2010)
Duties included: Build up and maintain *Drosophila melanogaster* stocks, prepare and conduct crosses of strains of drosophila containing different transposable elements and analyze the effects of different elements that have been inserted, analyze sequences of elements and predict possible effects elements have on altering genes near the insertion site, discuss and report findings and progress of experiment.

Teaching Experience

G848 Teachers Assistant (2/2018-5/2018)
Bioinformatics, Genomics, and Proteomics (Graduate Student Course)

Instructor, Bio 103L, (1/2013-5/2014)
California State University Dominguez Hills

Presentations/Publications

J. Victorino, M. Fox, W. Smith-Kinnaman, S. Peck, K. Hughes, A. Boyd, M. Zimmerly, R. Chan, G. Hunter, Y. Liu, AM. Mosley (2020). RNA Polymerase II CTD phosphatase Rtr1 fine-tunes transcription termination. PLOS Genetics

S. Peck, K. Hughes, **J. Victorino**, and A. Mosley (2019). Writing a Wrong: Coupled RNA Polymerase II Transcription and RNA Quality Control. WIREs RNA

T. Kunkle, S. Abdeen, N. Salim, AM. Ray, M. Stevens, A. J. Ambrose, **J. Victorino**, Y. Park, Q. Q. Hoang, E. Chapman, and S. M. Johnson (2018) Hydroxybiphenylamide GroEL/ES inhibitors are potent antibacterials against planktonic and biofilm forms of *Staphylococcus aureus*. Journal of Medicinal Chemistry

RB Slee, BR Grimes, R Bansal, J Gore, C Blackburn, L Brown, R Gasaway, JJeong**, J Victorino**, KL March, R Colombo, BS Herbert, and Murray Korc (2013). Selective inhibition of pancreatic ductal adenocarcinoma cell growth by the mitotic MPS1 kinase inhibitor, NMS-P715. Molecular Cancer Therapeutics

Rustbelt RNA Meeting Poster (10/2018)
Columbus, Ohio

IUSM Biochemistry and Molecular Biology Research Day Poster (10/30/2018)

Transcriptional Regulation by Chromatin and RNA Polymerase II Poster (10/04/2018)
(hosted by ASBMB), Snowbird, Utah

Purdue Chromatin and Epigenomics Symposium Poster	(06/2018)
Rustbelt RNA Meeting Talk Indianapolis, Indiana	(10/2017)
IUSM Biochemistry and Molecular Biology Research Day Talk	(09/2017)
IUSM Biochemistry and Molecular Biology Research Day Poster (Honorable mention)	(09/2017)
Chromatin and Epigenomics Symposium Talk	(10/2016)
CSUSH Student Research Day Oral presentation competition (2^{nd} Place)	(01/2013)
Summer Poster Symposium, Hosted by CRL at IUPUI	(07/2013)
Summer Poster Symposium, Hosted by CRL at IUPUI	(07/2012)

Professional Conferences/Courses

67^{th} American Society of Mass Spectrometry Annual Conference	(06/4/2019)
ASMS LC-MS: Practical Maintenance and Troubleshooting	(06/01/2019)
65^{th} American Society of Mass Spectrometry Annual Conference	(06/04/2017)
Mechanisms of Eukaryotic Transcription, Cold Spring Harbor, NY	(08/25/2015)
40^{th} SACNAS, San Antonio, TX	(10/03/2013)
25^{th} Annual CSU Biotechnology Symposium, Anaheim, CA	(01/03/2013)